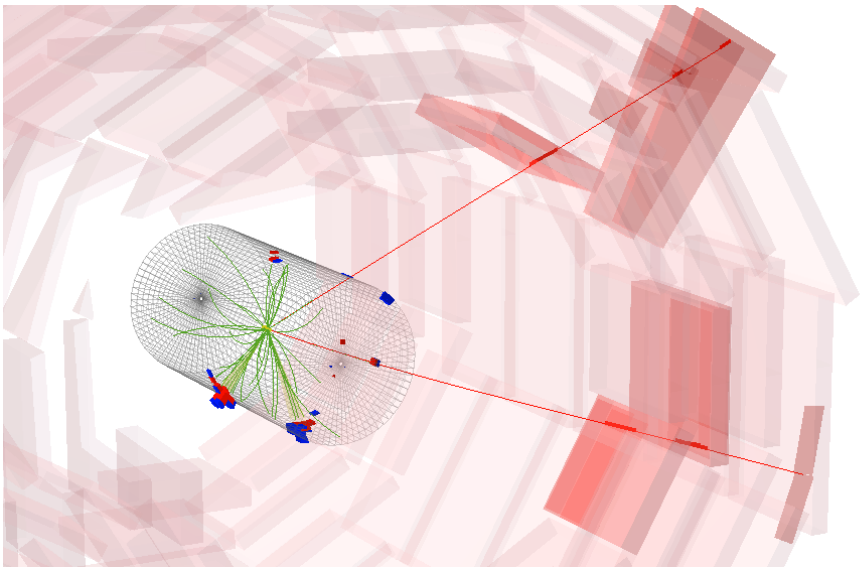
GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD
Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

Search for high mass Higgs bosons in the $\ell^+\ell^-q\bar{q}$ final state in CMS

Eduardo Navarro De Martino



CIEMAT, June 2015

Departamento de Investigación Básica
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

Departamento de Física Teórica
Facultad de Ciencias, Universidad Autónoma de Madrid

Search for high mass Higgs bosons in the $\ell^+\ell^-\mathrm{q}\bar{\mathrm{q}}$ final state in CMS

Búsqueda de bosones de Higgs de alta masa en el estado final $\ell^+\ell^-\mathrm{q}\bar{\mathrm{q}}$ en CMS

Thesis submitted by
Eduardo Navarro De Martino
for the degree of Doctor of Physics (Ph.D.)

Supervised by:
Dr. Pablo García Abia
Dr. José María Hernández Calama

Madrid, June 2015

Thesis Committee

Dr. Manuel Aguilar Benítez de Lugo (CIEMAT, Spain)
Prof. Jorge Fernández de Trocóniz Acha (U. Autónoma de Madrid, Spain)
Prof. María José Herrero Solans (U. Autónoma de Madrid, Spain)
Dr. Alexandre Nikitenko (Imperial College, UK)
Prof. Antonio Pich Zardoya (IFIC and Universidad de Valencia, Spain)

Dedicado a mi familia

Agradecimientos

Sirvan estas palabras de agradecimiento para todas las personas que han estado involucradas de una u otra manera en esta tesis doctoral, resultado de cuatro intensos años de trabajo. Mis disculpas por adelantado por las posibles omisiones.

En primer lugar, me gustaría agradecer a mis directores de tesis, Pablo y Chema (o Chema y Pablo, si se prefiere). Gracias por confiar en mí, compartiendo vuestra experiencia en beneficio de mi formación, no sólo en temas estrictamente científicos. Debo también reconcocer que ha sido un placer que hayáis sacado a relucir vuestro entusiasmo por la física durante estos años. Por último, me gustaría agradeceros que leyérais innumerables veces los primeros manuscritos, dando sabios consejos para ordenar mis ideas. De nuevo, gracias.

Un agradecimiento a mis colegas de la UAM que han colaborado fundamental y activamente en los estudios $2\ell 2q$ en CMS. A Jorge por su siempre acertado instinto en la física de partículas y todas las discusiones sobre el análisis. Gracias a Jorge y a Dermot por darle un empujón al análisis cuando parecía que se iba a quedar atrás.

También quiero mencionar a todos los integrantes del departamento de Investigación Básica, que me han dado la oportunidad de trabajar en el campo de la física de partículas experimental. Juntos habéis creado un entorno adecuado para desarrollar una investigación del calibre que se presenta en esta memoria. Quisiera reconocer a los *informáticos*, sin cuyo trabajo mi investigación no sería posible. Una agradecimiento especial a Antonio, por tener siempre razón y por la twiki, por ayudarme con las cosas de Grid, así como intentar recuperar aquellos datos que eran irrecuperables. Muchas gracias al grupo de CMS del CIEMAT por vuestros consejos, por vuestro interés o simplemente por estar ahí. Este párrafo no estaría completo si no agradeciese a los otros becarios de CMS, por sus muchas horas de discusiones, por quejarnos del ROOT, del CRAB y del combine, en resumen, por todos los buenos momentos: a Dani por ayudarme en mi etapa CERN, a Adrián por su sabiduría, el jazz y el rollerball (perdón, roller derby); gracias a Alberto por ser un alumno aplicado y por subir al despacho a discutir alguna idea que se te ocurría, como la búsqueda del *eduino* que seguro alguien ya te habrá robado. Albertismo, gracias

por apoyarme en los momentos en que lo necesitaba. A Mar (*majisma*) le deseo la mejor de las suertes con su tesis recién empezada y agradecerle su interés en mi última etapa de escritura. Tus dibujos no serán olvidados. Huelga decir que me hubiera gustado que compartiéramos más tiempo en tus estudios de top y de las DTs.

También quisiera agradecer al resto del grupo 2ℓ2q: a Óscar y Mathias por las tareas de supervisión y servir de enlace en la estructura de CMS, a Valentina por su paciencia y por no sucumbir a la presión del trabajo cuando más urgente era terminar los cálculos y a Kostya, que aunque estuvo poco, me enseñó cosas útiles.

Un agradecimiento igual de caluroso al resto de *becarios* del CIEMAT, que han conseguido que no perdiera la cordura durante estos años. Gracias a José, a Mariano (ese Betis), a Rafa y sus discusiones de ciencia ficción, a Jorge, a Cristina, a Carmen, a Isabel, a M.A., a Sara y a Manu grande, el viajero. Un reconocimiento también al Inge, becario honorífico. Gracias a todos por organizar las cenas de Navidad, por ir a *menjar tothom* y por los centenares de pastelitos.

Gracias a mis compañeros de la *beca-cueva*. Al gran Nacho por su fortaleza y por enseñarnos todos sus trucos. A Bruno por dejarnos meternos con él, por Monte, Carlo y por dejar de fumar; a Manuelinho por abrirnos su casa, por el tenisete y por la crista bendita; a Diana por amenizar el despacho con su constante buen humor y sus acertadas preguntas; y sobre todo a Javi, el *chico de las poesías*, por todos estos años codo con codo, por apoyarnos mutuamente en las dificultades y alegrías, por el cantar y por muchas cosas más que hemos vivido juntos.

Un reconocimiento especial a mi familia, que siempre me ha apoyado para llegar donde estoy hoy, han aguantado los días en los que llegaba muy tarde, cansado y de mal humor y han mostrado interés en mis viajes y otras aventuras. A mi hermano, disculparme por abandonarlo un poco por trabajar demasiado. A Daniel, la nueva incorporación, espero que lea este trabajo algún día y crezca admirando un poquito a su padrino.

Por último el agradecimiento más grande de todos es para Bárbara. Sin ti ni siquiera estaría escribiendo estas palabras. Gracias por ser el apoyo que siempre he necesitado, gracias por esperarme cuando has tenido que hacerlo, por estar en los malos momentos y también en los buenos, por no permitir que el tedio, la rutina o los golpes de la vida pudieran conmigo.

Contents

Foreword

The Standard Model of particle physics (SM) is a Quantum Field Theory that describes the elementary particles and their interactions, with one outstanding exception, gravity. It is an elegant theory, based on the fundamental symmetries of nature, that makes predictions of subatomic phenomena with unprecedented precision. Over the last 50 years its validity has been established experimentally with high precision in a wide energy range, up to the electroweak scale around 100 GeV.

In spite of its success, the SM is known to be an incomplete theory for various reasons. In the first place, it does not incorporate gravity and fails to accommodate dark energy and dark matter, evidenced in the last few decades by cosmological observations to be above 95% of the energy-matter content of the Universe. Other important deficiencies of the SM are the origin of mass, the strong CP problem, neutrino oscillations and matter-antimatter asymmetry. The fundamental nature of these phenomena firmly motivates the need for theories beyond the Standard Model (BSM). Current popular proposals include supersymmetry, string theory and extra dimensions, among others.

Predictions of these theories are explored in the Large Hadron Collider (LHC) at the electroweak scale and beyond. The LHC collides proton beams at unprecedented energies, 8 TeV at the proton-proton centre of mass up to year 2012, and 13 TeV since June 2015. This new energy regime has the potential to confirm the SM at the TeV scale and to discover new particles associated to unknown physics arising at higher energy.

The Higgs mechanism formulated in the 1960's was a convenient model to accommodate massive vector bosons in the SM, while preserving its renormalizability. The LHC collider and its experiments were designed with the discovery of the predicted Higgs boson (H) as one of their primary goals, being this particle a direct evidence for the realization of the Higgs mechanism in nature. Although no theoretical prediction for the mass of the Higgs boson existed in the SM, it was limited to be below the

TeV scale to prevent the amplitude of the W^+W^- scattering to exceed the unitarity bound in perturbation theory. Before the start of LHC, direct searches of the Higgs boson (at LEP and Tevatron) and indirect constraints from precision measurements of electroweak parameters (at LEP, SLD and Tevatron, among others) restricted the allowed range of Higgs mass values to be between 114.4 GeV and 160 GeV, and above 170 GeV, with lower values favored by the indirect measurements.

The work presented here is the analysis using CMS data of the search for Higgs bosons in the decay channel $H \to ZZ \to \ell^+\ell^- q\bar{q}$ where H decays into two Z bosons which subsequently decay to a lepton pair and a quark-antiquark pair, respectively. This analysis was originally designed to discover the SM Higgs boson in the high mass region, up to 1 TeV. After the discovery of the SM Higgs boson with a mass of 125 GeV by the ATLAS and CMS collaborations, this analysis contributed to establish the SM nature of the new boson by excluding additional SM Higgs bosons with masses in the range 200 GeV to 1 TeV, when combined with other H→ZZ and H→WW search analyses. At that point, the analysis was significantly improved and turned into a BSM Higgs boson search, probing a model with an additional electroweak singlet mixed with the SM Higgs boson. Failing to unveil signals of BSM Higgs bosons, with the limited number of events that CMS collected in Run I, we have been able to exclude ample regions of the parameter space in this model, which can be easily interpreted as limits on parameters of Two-Higgs-Doublet Models (2HDM).

Profiting from the advanced analysis techniques developed in this work, the analysis of larger data samples from Run II, already being collected by CMS at a centre-of-mass energy of 13 TeV, will allow to search for BSM Higgs boson signals and to reduce the number of potential BSM models.

Chapter 1

Introduction

The standard model of particle physics (SM) is the gauge quantum field theory that describes the constituents of matter and their interactions. The SM is in very good agreement with the high precision experiments performed during the last century. In the SM, the masses of the fundamental particles are acquired through the so-called Higgs mechanism, a spontaneous breaking of the electroweak gauge symmetry of the theory that predicts the existence of a massive scalar particle, the Higgs boson. This elementary particle had eluded detection during several decades, but has recently been discovered in proton-proton collisions produced at the Large Hadron Collider. Several theories extending the SM predict the existence of additional scalar particles with similar properties to those of the Higgs boson.

1.1 The mass mechanism in particle physics

The SM describes the interactions of elementary particles via the exchange of spin 1 intermediate bosons. They mediate three of the four known fundamental interactions: electromagnetic (photons, γ), strong (gluons, g) and weak (vector bosons, $W^{\pm}$ and Z) interactions. The theory classifies the constituents of matter, fermionic fields with spin 1/2, in three generations of two quarks (the up-type and down-type quarks) and two leptons (a charged lepton and its corresponding neutrino) with identical properties except their masses. The elementary particles included in the SM are summarized in Table 1.1.

The SM is gauge invariant under transformations of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry group. The $SU(3)_C$ gauge group is the fundamental symmetry of the quantum chromodynamics (QCD), the theory of the strong interaction [1,2] mediated by the exchange of eight massless gluons. Quarks and gluons carry colour, the quantum charge of the strong interaction. Due to the strength of α_s, the running coupling

Table 1.1: Elementary particles in the standard model of particle physics. q denotes de electric charge and m the mass.

Fermions			Bosons
up (u)	charm (c)	top (t)	gluon (g)
q(u)=2/3 m(u)≃2.3 MeV	q(c)=2/3 m(c)≃1.27 GeV	q(t)=2/3 m(t)≃173.2 GeV	q(g)=0 m(g)=0
down (d)	strange (s)	bottom (b)	photon (γ)
q(d)=-1/3 m(d)≃4.8 MeV	q(s)=-1/3 m(s)≃95 MeV	q(b)=-1/3 m(b)≃4.2 GeV	q(γ)=0 m(γ)=0
electron (e)	muon (μ)	tau (τ)	$W^{\pm}$ and Z
q(e)=−1 m(e)≃0.51 MeV	q(μ)=−1 m(μ)≃105.7 MeV	q(τ)=−1 m(τ)≃1776.8 MeV	q($W^{\pm}$)=±1; q(Z)=0 m($W^{\pm}$)≃80.38 GeV m(Z)≃91.19 GeV
electron neutrino (ν_e)	muon neutrino (ν_μ)	tau neutrino (ν_τ)	Higgs boson (h)
q(ν_e)=0 m(ν_e)<2 eV	q(ν_μ)=0 m(ν_μ)<0.19 MeV	q(ν_τ)=0 m(ν_τ)<18.2 MeV	q(h)=0 m(h)≃125.1 GeV

constant of QCD, quarks cannot be observed freely and form colourless compound states, consisting of three quarks or antiquarks (baryons), or quark-antiquark pairs (mesons). A quark or gluon produced in a high energy interaction hadronizes by combination with quark-antiquark pairs spontaneously created from the vacuum, producing a cascade of baryons and mesons.

The $SU(2)_L \times U(1)_Y$ gauge group describes the electroweak interaction, a unified picture of electromagnetic and weak interactions [3, 4]. It is mediated by four bosons, three related to the $SU(2)_L$ group (W_1, W_2, W_3) and the other one (B_0) to the $U(1)_Y$ transformations. The renormalizability of the theory imposes that the electroweak mediators are massless bosons [5]. However, it is well known that the gauge weak vector bosons are massive, confirmed experimentally by the measurements of the W and Z masses [6, 7]. This contradiction is solved by spontaneously breaking the $SU(2)_L \times U(1)_Y$ symmetry. The so-called electroweak symmetry breaking mechanism [8–12] adds a new component to the SM, Φ, a $SU(2)_L$complex scalar field with four real components, ϕ_i, the so-called the Higgs field:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{1.1}$$

Its lagrangian density is:

$$L_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \ , \tag{1.2}$$

where:

$$V(\Phi) = -\mu^2 |\Phi|^2 + \lambda |\Phi|^4 \ . \tag{1.3}$$

Here the units are chosen so that $\hbar = c = 1$ and μ and λ are the mass and self-interaction coupling constant of the scalar field. The Higgs lagrangian is, by construction, invariant under $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ transformations. However, the symmetry is spontaneously broken by a non-zero vacuum expectation value of the Higgs field:

$$\langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \ , \tag{1.4}$$

with $v = \sqrt{\mu^2/\lambda}$, while preserving the invariance under $\mathrm{U(1)_{EM}}$ transformations, carrying electric charge. As a consequence, the W_i and B_0 massless particles are mixed, resulting into the massive vector gauge bosons $\mathrm{W}^\pm$ and Z, and the massless photon. The masses of the $\mathrm{W}^\pm$ and Z bosons, $m_{\mathrm{W}^\pm}$ and m_{Z}, are acquired by absorbing three degrees of freedom of the Φ doublet into the longitudinal components of the physical vector bosons. The masses of the gauge bosons are proportional to v and to the gauge couplings of the $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$ groups, g and g':

$$m_{\mathrm{W}^\pm} = \frac{vg}{2} \qquad\qquad m_{\mathrm{Z}} = \frac{v\sqrt{g^2 + g'^2}}{2}. \tag{1.5}$$

The remaining degree of freedom of Φ manifests itself as a new scalar elementary particle, the Higgs boson, h.

The electroweak symmetry breaking mechanism also gives mass to the elementary fermions by the inclusion of the Yukawa lagrangian density. The masses of the fermions arise from interactions with the Higgs field:

$$m_f = \frac{vY_f}{\sqrt{2}} \ , \tag{1.6}$$

where Y_f, the Yukawa couplings, are measured experimentally.

The value of the Higgs boson mass, m_{h}, is related to the parameters of the Higgs potential (Equation 1.3):

$$m_{\mathrm{h}} = v \cdot \sqrt{2\lambda} \ . \tag{1.7}$$

The vacuum expectation value of the Higgs field is related to the mass of the vector bosons (Equation 1.5), and it is inferred from the measurements of $m_{\mathrm{W}^\pm}$ and m_{Z} to be approximately 246 GeV [13]. However, the self-coupling of the Higgs boson, λ,

can only be determined by direct production and detection of Higgs bosons. Therefore, the mass of the Higgs boson is not predicted by the SM and the experiments aiming to detect the predicted new scalar are designed to explore a wide range of possible values of m_{h}.

Before the start of the Large Hadron Collider (LHC), the most sensitive direct search was performed at the LEP collider, excluding the presence of a Higgs boson with masses below 114.4 GeV[(1)] [14]. The high precision measurements of the $\mathrm{W}^{\pm}$ and Z bosons properties and the electroweak radiative corrections performed by the LEP and SLD experiments placed indirect constraints on m_{h}, below 168 GeV [15].

In 2012, the ATLAS and CMS experiments reported the observation of a new boson [16, 17] with a mass around 125 GeV and properties consistent with those predicted for the SM Higgs boson, h(125). After the discovery of the SM Higgs boson, theories predicting an extended scalar sector yielding additional Higgs bosons gained importance. In particular, because this new scalar sector could be connected to dark matter [18] and, most importantly, because it could be accessible in the high energy proton-proton collisions at the LHC. The existence of a new scalar sector containing an additional Higgs-like state is explored in this work.

One of the simplest theories beyond the SM (BSM) involving an enlarged scalar sector is the electroweak singlet extension of the SM [19–21]. This model allows to modify the SM predictions for a heavy Higgs boson and includes the possibility of non-SM Higgs boson decays. In this extension of the SM, a new field, Φ_S, scalar under $\mathrm{SU(2)_L}$ transformations, modifies the lagrangian of the scalar sector given in Equation 1.2:

$$L_{\mathrm{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) + (D_\mu \Phi_S)^\dagger (D^\mu \Phi_S) - V(\Phi, \Phi_S) \ , \tag{1.8}$$

where the Higgs potential of Equation 1.3 includes the new scalar field:

$$V(\Phi, \Phi_S) = -\mu^2 |\Phi|^2 + \lambda |\Phi|^4 - \mu_S^2 |\Phi_S|^2 + \lambda_S |\Phi_S|^4 + \eta |\Phi|^2 |\Phi_S|^2 \ . \tag{1.9}$$

In the unitary gauge:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \phi + v \end{pmatrix} \ , \quad \Phi_S = \frac{1}{\sqrt{2}} \left(\phi_S + \xi \right) \ . \tag{1.10}$$

The vacuum expectation value of the Higgs singlet field, ξ is:

$$\langle 0 | \Phi_S | 0 \rangle = \frac{\xi}{\sqrt{2}} = \sqrt{\mu_S^2 / 2\lambda_S} \ . \tag{1.11}$$

[(1)] In this work natural units are assumed, where the Planck constant and the speed of light are equal to one ($\hbar = c = 1$). In that context, energies, masses and momenta of particles are measured using the same units, multiples of eV.

The two remaining degrees of freedom are mixed in the last term of the potential, determined by the η parameter. The ϕ degree of freedom can be identified with a SM Higgs boson, while ϕ_S is related to the new sector. Due to the mixing of these states, their mass matrix is not diagonal. The mass eigenstates are obtained diagonalizing that matrix, resulting in two physical scalar particles: the Higgs boson, h, and the so-called electroweak singlet particle, $\mathrm{H_S}$, where:

$$\mathrm{h} = \quad \cos\omega \cdot \phi + \sin\omega \cdot \phi_S \tag{1.12a}$$

$$\mathrm{H_S} = -\sin\omega \cdot \phi + \cos\omega \cdot \phi_S \tag{1.12b}$$

The C and C' parameters are defined as a function of the mixing angle, ω, as:

$$C = \cos\omega \ , \quad C' = \sin\omega \ , \tag{1.13}$$

which satisfy the relation $C^2 + C'^2 = 1$.

As a consequence of the mixing of both states, the h production cross section, σ, and total width, Γ, are modified with respect to the SM predictions, σ_{SM} and Γ_{SM} respectively:

$$\sigma = C^2 \cdot \sigma_{SM} \ , \tag{1.14a}$$

$$\Gamma = C^2 \cdot \Gamma_{SM} \ . \tag{1.14b}$$

Similarly, the cross section σ' and width Γ' of the electroweak singlet particle is related to the SM prediction:

$$\sigma' = C'^2 \cdot (1 - B_{new})\sigma_{SM} \ , \tag{1.15a}$$

$$\Gamma' = \frac{C'^2}{1 - B_{new}}\Gamma_{SM} \ , \tag{1.15b}$$

where B_{new} denotes the decay branching ratio of the electroweak singlet to channels not allowed for the SM Higgs boson, such as $\mathrm{H_S} \rightarrow$h(125)h(125).

1.2 Production and decay of a Higgs boson in hadron colliders

A SM scalar boson can be produced in hadron colliders mainly via four different mechanisms (Figure 1.1). The dominant process up to the TeV scale is the gluon-gluon fusion mechanism (ggH), or gluon fusion for short, where the interaction of two gluons from the colliding protons produces a Higgs boson via a virtual loop of

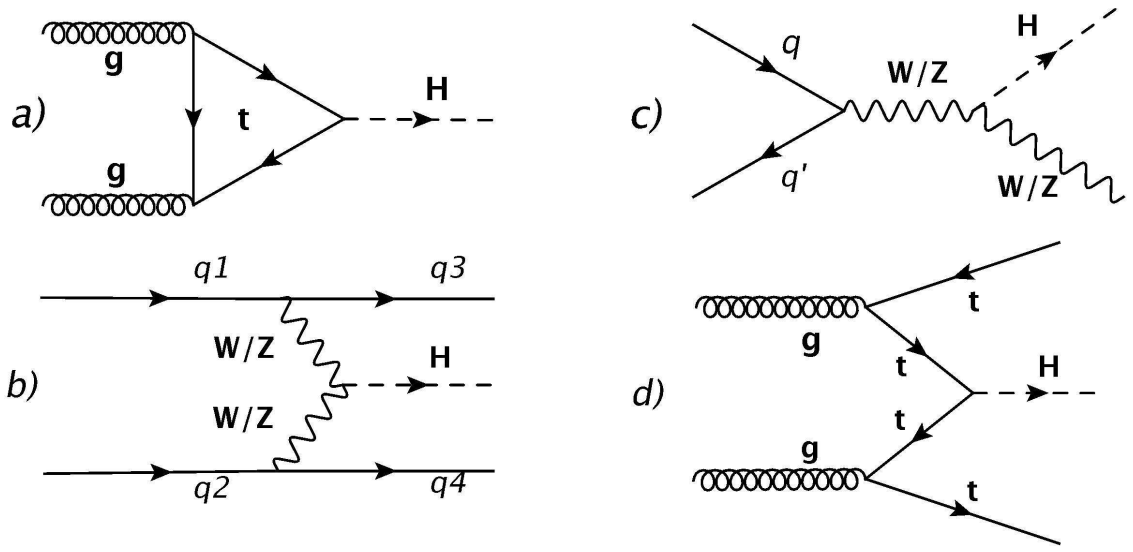


Figure 1.1: Lowest order Feynman diagrams for the main Higgs boson production mechanisms at the LHC: (a) gluon-gluon fusion (ggH), (b) vector boson fusion (VBF), (c) Higgs-strahlung (VH) and (d) H production in association with $t\bar{t}$ (ttH).

heavy quarks (predominantly top). In the vector boson fusion production mechanism (VBF), the Higgs boson is produced by the interaction of two vector bosons emitted by the quarks of the protons. This process becomes especially relevant for Higgs bosons with very high masses, where its cross section is comparable to the ggH cross section. The other two production mechanisms have lower cross sections and are only relevant for low Higgs boson masses. They are the Higgs boson production either in association with a vector boson (Higgs-strahlung, VH) or in association with a top quark pair (ttH). The Higgs boson production cross section in proton-proton interactions depends on its mass and the centre-of-mass energy of the collisions (Figure 1.2).

The natural width of the Higgs boson depends on m_{h}(Figure 1.3). Its value is a few MeV at low mass and rapidly increases up to almost 1 TeV at very high m_{h}. The large total width implies a quick decay into other elementary particles, with lifetimes varying from 10^{-20} to 10^{-25} s. Hence, the Higgs boson can only be detected through its decay products.

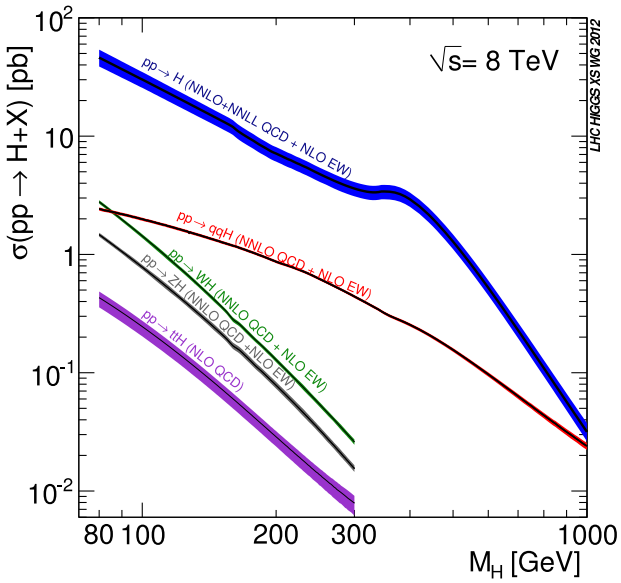


Figure 1.2: Cross sections for the main H production mechanisms, as function of the Higgs boson mass, M_H, for proton-proton collisions at $\sqrt{s} = 8$ TeV [22].

The Higgs boson primarily decays to the heaviest pair of kinematically available particles. The decay branching fractions to different final states are shown in Figure 1.4 (left) for a Higgs boson mass range between 80 and 1000 GeV. Low mass Higgs bosons, up to 150 GeV, decay mainly into fermions, predominantly in $b\bar{b}$ pairs. At higher masses, the decays $H\rightarrow W^+W^-$ and $H\rightarrow ZZ$ become dominant and above the top quark pair mass threshold ($m_H > 2m_t$) the contribution of the decay to $t\bar{t}$ is significant. The Higgs boson can also decay into a pair of massless bosons, such as gluons or photons, through loop diagrams involving heavy particles as the W boson or the top quark. These decays are only relevant in the low mass region.

The sensitivity of each decay channel to a Higgs boson signal varies significantly with the Higgs boson mass hypothesis. The LHC experiments ATLAS and CMS perform direct searches exploring all the final states ($H\rightarrow b\bar{b}$, $H\rightarrow \tau\tau$, $H\rightarrow ZZ$, $H\rightarrow WW$, $H\rightarrow\gamma\gamma$, $H\rightarrow Z\gamma$, $H\rightarrow\mu\mu$) in the range from 114 GeV to 1 TeV. The first observation of the Higgs boson [16,17] is the result of the extensive search these experiments performed, with its measured properties to be consistent with the SM predictions [24–27]. The measured mass of the Higgs boson, m_h=125.09±0.21(stat.)±0.11(syst.) GeV, is obtained using the $H\rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = \mu$, e) and $H\rightarrow\gamma\gamma$ decay channels [28]. Despite their lower branching fractions they have the highest sensitivity to a low mass Higgs boson due to their clean signature and excellent mass resolution.

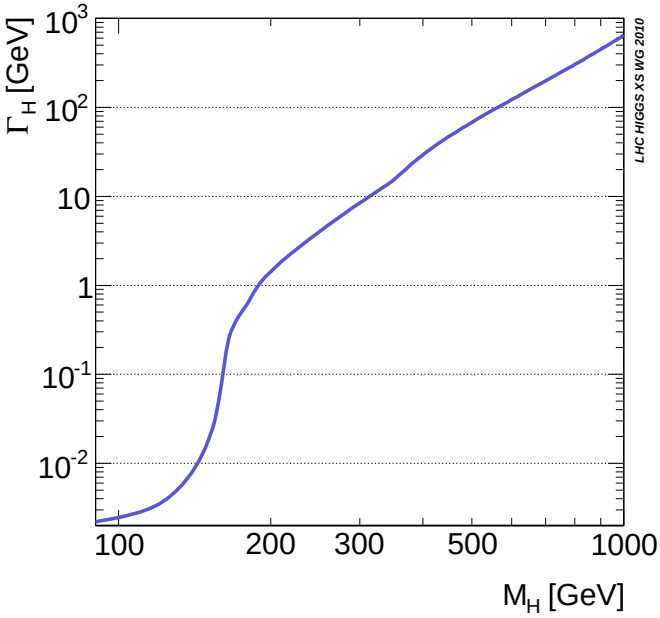


Figure 1.3: Total width of the SM Higgs boson, Γ_H, as a function of its mass, M_H [22].

1.3 The H → ZZ →$\ell^+\ell^-$q$\bar{\text{q}}$ decay channel

After the discovery of the SM Higgs boson, h(125), the possibility of finding more particles on an extended scalar sector gains interest, as predicted by many BSM theories. In this context, it is extremely important to search for high mass Higgs bosons, combining all the relevant channels available to reach the highest sensitivity. The absence of a new scalar particle would be an additional measurement supporting that h(125) is the Higgs boson predicted by the SM.

One of the most sensitive channels to hypothetical heavy Higgs boson signals, H, is the H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ decay. Its main advantage is that all the leading particles of the event can be detected, allowing a full kinematic reconstruction and, therefore, providing a powerful discrimination against a potentially huge background of SM processes. This channel has a large branching ratio, mainly due to the hadronic decay of the Z boson, $B(\text{Z}\rightarrow \text{q}\bar{\text{q}}) = 69.91\%$ [13]. Moreover, it also profits from the pair of leptons coming from the leptonic Z decay, Z→$\ell^+\ell^-$. The presence of a high momentum dilepton pair spatially separated from the hadronic jets provides a clean signature that can be used to efficiently trigger the data acquisition system and to further isolate the signal.

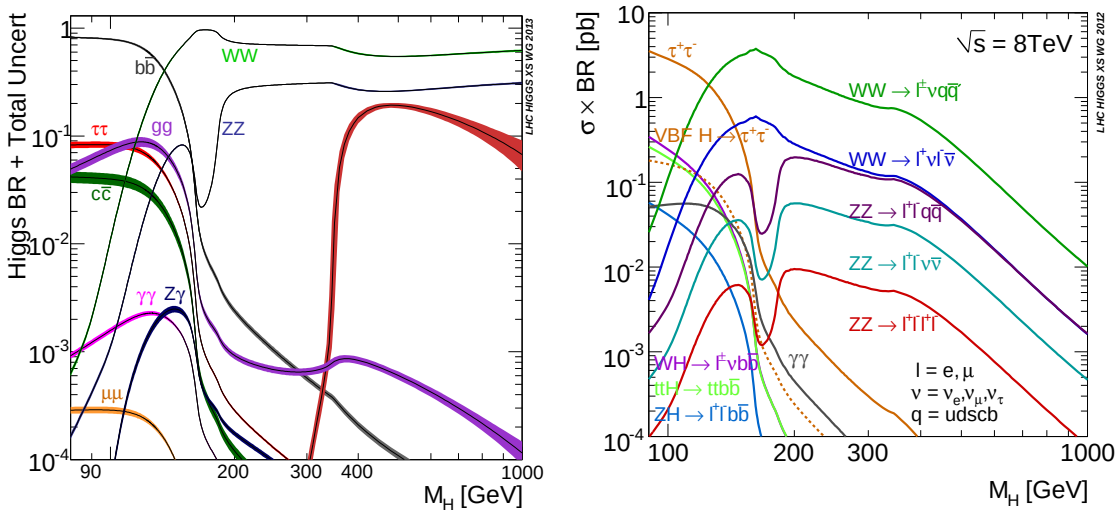


Figure 1.4: Left: decay branching fractions of a SM Higgs boson as function of its mass, M_H [22]. The shaded areas represent the theoretical uncertainties. Right: production cross section times branching fraction of a SM Higgs boson as function of its mass, M_H, for the main decay channels considered by the CMS experiment [23].

The production of a Higgs boson is an extremely rare process. The right plot of Figure 1.4 depicts the cross section times branching fraction of the diboson decay channels used by the ATLAS and CMS collaborations to search for a Higgs boson with masses between 80 and 1000 GeV. The $H \to ZZ \to \ell^+\ell^- q\bar{q}$ process has the highest branching ratio among the ZZ decays considered in CMS ($\sigma \times B \sim$1-100 fb). At the energy of the LHC, the cross section of the inelastic proton-proton collisions is of the order of tens of mb, about ten orders of magnitude higher than the predicted heavy Higgs boson production cross sections, of the order of 10^{-2} to 1 pb depending on its mass. Furthermore, several SM processes with higher cross sections and the same signature than $H \to ZZ \to \ell^+\ell^- q\bar{q}$ constitute a serious background source. The dominant contribution comes from the Drell-Yan Z boson production in association with jets (Z+jets), which has a cross section from 10^3 to 10^5 times greater than the H production cross sections. Another important source of background arises from top-quark pair production, following the decay chain $t\bar{t} \to 2\ell 2\nu b\bar{b}$, having a cross section of hundreds of pb. Finally, another small but relevant source of background is the electroweak vector boson pair production (ZZ, WZ and WW), with cross sections of the order of tens of pb. A precise knowledge of these backgrounds is an essential key in the search for heavy Higgs bosons.

◇ ◇ ◇

This thesis is devoted to the search for a high mass Higgs boson with H→ZZ decay and with subsequent decays Z→$\ell^+\ell^-$and Z→q$\bar{\text{q}}$. The analysis presented in this work was originally designed to cover the high mass region of the SM Higgs boson, above 200 GeV up to 1 TeV, in the context of the extensive search using all available decay channels. After the h(125) discovery, the results of the analysis presented here are used to firmly establish that the observed Higgs boson is consistent with the SM predictions, by excluding additional SM Higgs boson signals. The analysis is also interpreted in the context of a BSM search for an additional scalar particle, probing the electroweak singlet model.

The next chapters (2 and 3) briefly describe the CMS detector, the data used in the analysis and the lepton and quark reconstruction. The details of the analysis and the interpretation of the results are given in chapters 4 and 5.

Chapter 2

The CMS experiment at the LHC

Particle physics up to the TeV energy scale is explored in proton-proton collisions at the Large Hadron Collider. The particles resulting from the colliding proton beams are collected by the CMS experiment in order to study the validity of the SM and to search for new phenomena. This chapter describes the LHC accelerator and the CMS detector.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the most powerful particle accelerator in the world, located in the French-Swiss border, near the CERN laboratory in Geneva [29]. LHC is a proton-proton collider, composed by two circular rings with a total length of 27 km, built in the 100 m underground tunnel originally constructed to house LEP, the Large Electron-Positron collider.

The LHC is designed to operate at unprecedented energy and intensity in high energy physics, colliding two 10^{14}-proton beams at a centre-of-mass energy of $\sqrt{s} = 14$ TeV. These beams circulate in opposite directions on each of the LHC rings, bent by 1232 superconducting magnets cooled down to 1.9 K. The beams are focused using quadrupole magnets and intersected in four collision points, where the four main detectors, ATLAS, CMS, ALICE and LHCb are located (Figure 2.1). ATLAS [30] and CMS [31] are general-purpose detectors designed to withstand the high-luminosity of proton-proton collisions provided by the LHC. Their physics goals include the study of a large variety of phenomena, such as QCD, electroweak and flavour physics, and to perform searches for new particles. ALICE [32] is designed to study lead ion collisions (p-Pb and Pb-Pb) and LHCb [33] is specialized in detailed studies of b quark physics and CP-symmetry violation. Both ALICE and LHCb are exposed to lower luminosities than ATLAS and CMS.

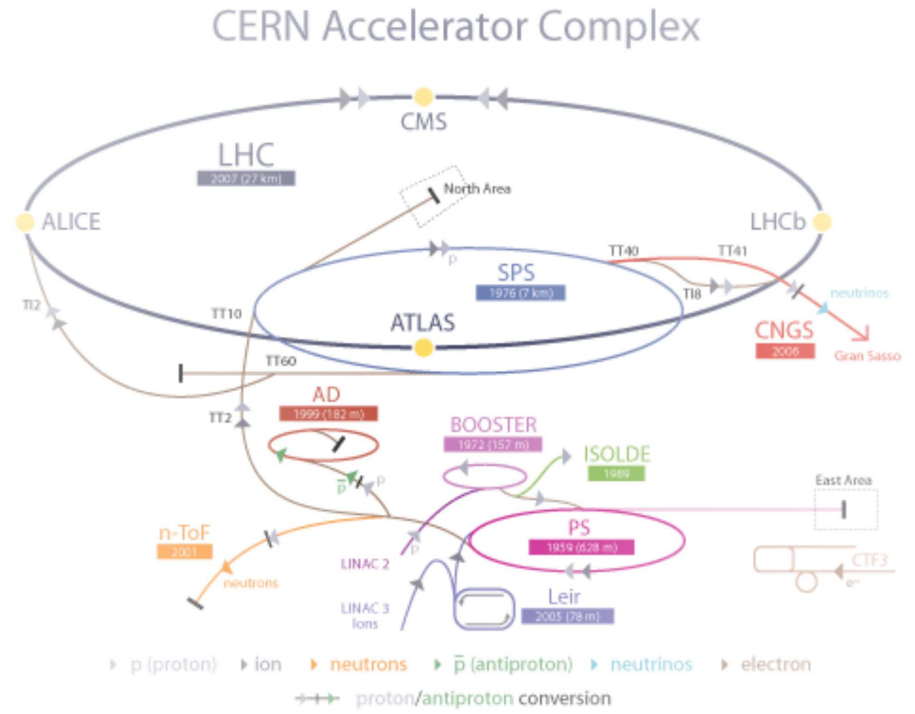


Figure 2.1: Upper: the LHC accelerator. Lower: LHC is the largest of the CERN accelerator complex. The experiments are marked by yellow bullets. CMS is on top of the drawing, opposite to ATLAS and CERN main site.

Physical processes of interest for the LHC experiments involve the production of heavy particles, instantaneously decaying to high transverse momentum secondary particles. The low cross section of these processes results in a small number of events, N. To produce such rare events the high energy proton beams are focused and collided at high rates, crossing each other every 50 ns. The number of protons crossing a unit surface (cm^2) per unit time (s) is the so-called instantaneous luminosity, $\mathcal{L}$. The number of events produced per unit time is:

$$\frac{d\mathrm{N}}{dt} = \mathcal{L} \cdot \sigma \ . \tag{2.1}$$

The total number of events is $N = L \cdot \sigma$, where $L = \int \mathcal{L} \cdot dt$ is the luminosity integrated during the data taking period. At the LHC, with a design value of $\mathcal{L}$=10^{34} $cm^{-2}s^{-1}$, the inelastic proton-proton cross section of $\sim$ 100 mb gives event rates of 10^9 events/s or 10^{15} events/year. This creates a busy environment for the high momentum physics of interest, which has cross sections of the order of the pb or less, typically with about 10^{12} times lower rates (10^{-2} to 10^{-4} events/s or 10 to 10^3 events/year).

In the Run I of the LHC, proton beams were collided at centre-of-mass energies, $\sqrt{s}$, of 7 TeV (year 2011) and 8 TeV (year 2012), well below the design energy of 14 TeV. Most of the Run I integrated luminosity, around 20 fb^{-1}, was delivered during the 8 TeV run. The data collected by the CMS experiment in this period are used in the analysis reported here, with peak instantaneous luminosities as high as 7.7×10^{33} $cm^{-2}s^{-1}$.

At the extreme luminosities reached by the LHC, multiple proton-proton interactions occur per bunch crossing. This effect, known as pile-up (PU), plays an important role in the analyses performed by the LHC experiments. In the 8 TeV run, the number of pile-up events ranged from 15 to 30 interactions per bunch crossing, being 20 on average.

2.2 The CMS detector

The Compact Muon Solenoid experiment [31] (CMS) is a general purpose detector, aiming to perform physics studies from the analysis of proton-proton collisions delivered by the LHC. The main objective of CMS is to search for the SM Higgs boson in a wide range of Higgs boson masses, from about 100 GeV up to 1000 GeV and once it is found, to characterize its properties (mass, spin, parity, width and couplings). CMS will also search for new particles arising in BSM theories, such

as dark matter, supersymmetric particles, additional vector bosons or more scalars predicted by extended Higgs sectors. Generally, these models predict the existence of new physics at the scale of the TeV. An important part of the CMS physics program is to perform high precision studies of SM physics: QCD, electroweak and flavour physics, that provide complementary information with respect to the direct searches, since any deviation from the SM prediction would be a hint of new physics. Last but not least, CMS will search for a large variety of exotic physics, which include extra dimensions, microscopic black holes, lepton flavour violation or heavy quarks, among others.

CMS consists of a cylindrical barrel of 21.6 m length, 14.6 m diameter, closed by two endcaps. The different subdetectors that conform the CMS apparatus are situated concentrically, centred at the collision point (Figure 2.2). The detector is designed to be highly hermetic, capturing the particles emerging from the proton-proton collisions in almost a 4π solid angle around the interaction point, with a highly efficient data acquisition system at a collision rate of 40 MHz, equivalent to instantaneous luminosities of 10^{34} cm^{-2}s^{-1}. The purpose of the CMS subdetectors is to measure with high resolution the energy, momentum and position of electrons, muons, photons and hadrons (jets). The neutrinos can be inferred thanks to the CMS hermeticity, through the transverse-momentum imbalance of all reconstructed visible particles in the event, the so-called missing transverse energy, $\not\!\!E_T$.

In the coordinate system adopted by CMS, the origin is set at the interaction point in the centre of the barrel, the y-axis points vertically upward, the x-axis points radially inward to the centre of the LHC rings and the z-axis points along the beam direction, anticlockwise. In cylindrical coordinates, the azimuthal angle ϕ is measured from the x-axis in the transverse (x-y) plane, while the polar angle θ is measured from the z-axis. Instead of the θ coordinate, the pseudorapidity variable is often used[(1)]:

$$\eta \equiv -ln\left[tan\left(\frac{\theta}{2}\right)\right] \ . \tag{2.2}$$

[(1)] η is zero in the x-y plane and $\pm\infty$ in the beam axis, $\eta = 1$ corresponds to $\theta \simeq 40°$ and $\eta = 2.5$ to $10°$.

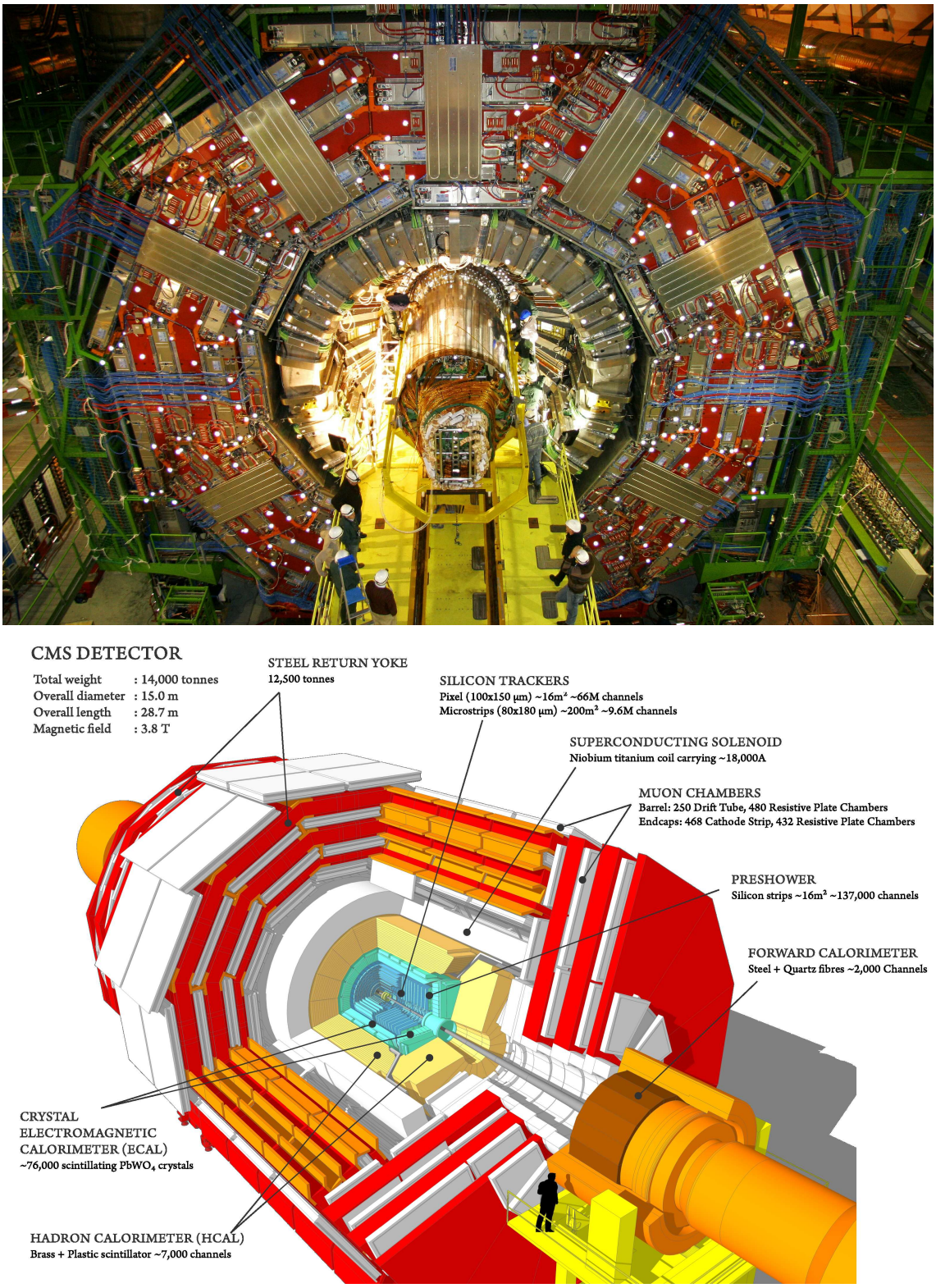


Figure 2.2: Upper: side view of the CMS detector (x-y plane) during the insertion of the silicon tracker. The DT chambers built at CIEMAT are visible in the outer layers of the detector. Lower: Schematic of the CMS apparatus and its different subdetectors [34].

The central feature of the CMS detector is a superconducting solenoid of 12.5 m length and 6 m internal radius [35], which provides a uniform magnetic field of 3.8 T in the +z direction. Outside the solenoid, the magnetic lines return through an iron yoke. The intense magnetic field allows a precise measurement of the transverse momentum, p_T, of the relativistic charged particles using the curvature of their trajectories. Within the superconducting solenoid volume a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections, are placed. Embedded in the steel yoke outside the solenoid, gas-ionization chambers detect the muons.

The innermost subdetector is the tracking system [36]. The goal of the tracker is to reconstruct the trajectories of the charged particles and the vertices where the interactions occur. This system is composed by two detectors: the pixel tracker, made of 66 million pixels, and the silicon microstrip tracker, composed by 9.3 million strips. The pseudorapidity coverage of the tracker system is $|\eta| < 2.5$. The high segmentation of this subdetector is crucial for the track reconstruction in a high density environment. The track detection efficiency is very high, around 99% for charged particles with transverse momentum greater than 100 GeV up to $|\eta|$=2 [34]. The tracker system also provides a robust vertex reconstruction. The transverse impact parameter[(2)] resolution for particles with p_T >10 GeV is better than 35 μm, while the longitudinal impact parameter resolution is better than 75 μm [36]. The momentum resolution in the tracker is better than 3% in the central part of the detector ($|\eta| < 1.5$) for particles with transverse momentum up to 100 GeV. The jet momentum is determined as the vectorial sum of all particle momenta inside the jet. Comparisons with simulations determined that the jet transverse momentum resolution is between 5% and 10% over the whole p_T spectrum and detector acceptance.

The spatial resolution of the tracker permits to distinguish between tracks coming from the hard proton-proton interaction and from the pile-up. Moreover, the tracker performance allows to identify jets containing a bottom quark, using b-tagging algorithms. The identification of b jets is interesting for several studies: top quark decays via the t→bW process, with a branching fraction of almost 100%; many particles predicted by BSM theories have final states involving bottom quarks; it allows to study flavour and CP-violation physics. In the analysis presented in this thesis the b-tagging is one of the most powerful methods to enhance the signal against the overwhelming background contribution.

The ECAL system [37] measures the energy and direction of electrons and photons,

[(2)]The transverse (longitudinal) impact parameter is defined as the minimum transverse (longitudinal) distance of the track to the interaction point.

and part of the energy of hadrons and muons. This subdetector is made of almost 76000 $PbWO_4$ scintillating crystals, covering $|\eta| < 3$. Electrons and photons entering the ECAL generate electromagnetic cascades, absorbed inside the crystals and measured using photodetectors. The high density of the crystals produces a short Molière radius which together with its thin segmentation results in a fine granularity of the lateral shape of the showers. The photon energy resolution with transverse energy around 60 GeV varies between 1.1% and 2.6% over the solid angle of the ECAL barrel, and from 2.2% to 5% in the endcaps. The energy resolution for electrons with transverse energy around 45 GeV is better than 2% in the ECAL barrel central region ($|\eta| < 0.8$), and is between 2% and 5% in the other regions [38].

The HCAL system [39] completely absorbs hadronic jets, measuring its energy and direction. The HCAL Barrel and HCAL Endcaps surround the ECAL with a coverage up to $|\eta| < 3$, located inside the magnet coil. Hence, the HCAL is constructed using non-magnetic material, brass plates interlayed with plastic scintillation tiles. To fully contain the hadronic showers in the barrel, an additional array of scintillation tiles are placed outside the magnet: the HCAL Outer. Full coverage is provided in the forward region ($3 < |\eta| < 5$) by the HCAL Forward, made of quartz fibres that emit Cerenkov light, collected by a copper absorber. The HCAL system is able to reconstruct jet energy with a resolution of 15% for particles with p_T=10 GeV, 8% for p_T=100 GeV, and 4% for p_T=1 TeV.

Muons are the only interacting particles that are not completely absorbed in the electromagnetic or hadronic calorimeters. The muon system [40] is located outside the solenoid, with three different types of gas-ionization chambers. In the barrel region ($|\eta| < 1.2$) drift tube chambers (DT) are located, where the expected rate of muons and residual magnetic field is low. The Cathode Strip Chambers (CSC) are situated in the endcaps ($0.9 < |\eta| < 2.4$) due to its working capabilities under high particle rate and magnetic field conditions. Finally, Resistive Plate Chambers (RPC) are interlocated between the barrel and endcap regions ($|\eta| < 1.6$), with the goal of triggering muon events. Muon tracks detected in the muon system (matched to tracks measured in the tracker system) result in relative transverse momentum resolutions of 1.3%–2% in the barrel and better than 6% in the endcaps, for muons with $20 < p_T < 100$ GeV. The p_T resolution of muons with p_T=1 TeV in the barrel is better than 10% [41].

CIEMAT participated in the design and construction of the CMS muon detector, building 30% of the DT chambers and the associated electronics. CIEMAT is actively involved in the trigger, alignment, online/offline monitoring and in the muon reconstruction software.

CMS performs a clear and efficient identification of the particles emerging from the proton collisions. Each type of elementary particle reaching the detector imprints a different signal. Electrons are charged particles, leaving a track in the tracker system and depositing their energy in the ECAL system. Photons are neutral particles, depositing their energy in the ECAL, without an associated track in the tracker system. Hadrons deposit most of their energy in the HCAL, with minimum interaction in the ECAL. Charged hadrons also leave a trace in the tracker, while neutral hadrons only interact with the calorimeters. Muons are charged particles, leaving a trail in the tracker and in the muon systems, with minimum interaction in both calorimeters.

The data acquisition, trigger and processing systems

The typical size of a CMS event is around one MB (1024^2 bytes). The high collision rate of 20 MHz in Run I leads to 5×10^8 events/s, which implies a storage rate of 1 PB/s. In order to reduce the data collection rate to a manageable level, of few hundred MB/s, the trigger system [42] selects in real time the most relevant events to be recorded for their posterior analysis. It consists of a series of algorithms, designed to keep potentially interesting high-momentum events, while discarding low energy processes. The CMS trigger system is implemented in two sequential levels: the level-1 (L1) and the high-level trigger (HLT).

The L1 trigger system is hardware-based, built using custom electronics installed near the detector front-end electronics. It performs a quick decision to filter each event within 3 μs. Due to the time limitation, it only uses simple information from the ECAL, HCAL and muon systems, without information of the tracker, and applies fast algorithms using pattern recognition and a fast summing technique. The L1 trigger reduces the rate from 20 MHz to 100 kHz.

Events passing the L1 filter enter the HLT system, a set of complex event filtering algorithms that run in a CPU farm of about 10^4 cores built from commodity hardware. The HLT system has access to data from all the CMS subdetectors with full granularity and resolution. The average decision time to filter an event in the HLT is around 100 ms. The reduction rate is of the order of 10^3, with an output rate of few hundred Hz. To optimize the data flow, the HLT first uses information of a partial event reconstruction, using directly detector signals from the calorimeters and the muon systems. If the event is not discarded, the full information is used to make the final decision. The HLT also classifies the events in about 50 primary datasets attending to the objects passing the trigger. For instance, events used for

the analysis presented in this work are triggered by the presence of two high momentum leptons with 17 GeV and 8 GeV p_T-thresholds, classified in the double-muon, double-electron or muon-electron samples.

After the trigger step, achieving a rejection factor of nearly 10^5 (resulting in a storage rate of a few PB/year), data are stored, reconstructed and analyzed by the CMS offline computing system [43]. This system is a distributed infrastructure that interconnects resources in geographically separated locations by means of high throughput Internet networks, the World-wide LHC Computing Grid (WLCG). WLCG integrates 3×10^5 CPU cores, 400 PB storage capacity and provides computing resources to about 10^4 users.

The Grid is layered in a tiered structure with three levels. The Tier-0 is a CPU farm based at CERN, in charge of the prompt processing of the raw data coming from the detector to create datasets containing the reconstructed physical objects. It also performs fast calibration and monitoring tasks. The Tier-1s are 11 computing centres around the world intended to store the raw and processed data and to perform reconstruction, calibration and other data-intensive tasks. They provide skimmed datasets containing the essential information for the analyzers, with an important reduction in number of events and size. About 50 Tier-2 centres, based in the institutes of the CMS collaboration, provide data storage and CPU power to perform simulations and data analysis.

CIEMAT contributes to the WLCG infrastructure with a Tier-1 centre, PIC, operated together with IFAE in Barcelona, and a Tier-2 centre in Madrid. Both sites provide 5% of the compute and storage resources of WLCG. CIEMAT has been very actively involved since its inception in the development, integration and operation of the WLCG distributed computing system.

Simulations based on Monte Carlo techniques are used to generate events as similar as possible to those produced in real collisions, using a Geant-4 [44] simulation of the CMS detector response. These simulated events are processed with the same software as the data collected by CMS. They are used to test the detector performance (resolution, efficiency) and to study the physics of the processes that would be produced by the collisions (topology, kinematics). In general, the size of the simulated samples is around ten times the size of the reconstructed samples, in order to reduce the statistical fluctuations when comparing to the data.

◇ ◇ ◇

Since the year 2010, the LHC has collided proton beams with centre-of-mass energies of 7 TeV (2010 and 2011) and 8 TeV (2012). The CMS experiment collected a data sample corresponding to an integrated luminosity of almost 25 fb^{-1} (5 fb^{-1} in 2011 and 20 fb^{-1} in 2012). The analysis of the events consistent with $H \to ZZ \to \ell^+\ell^- q\bar{q}$ final states presented in the next chapters is a good example of the excellent performance of the CMS detector and the high quality of the data it collected.

Chapter 3

Reconstruction of the $\ell^+\ell^-\mathrm{q\bar{q}}$ final state

The analysis of the data collected by CMS, used to search for high mass Higgs bosons, is based on the kinematic information of the events, which is extracted from the electronic signals recorded in the detector. This so-called *event reconstruction* provides *physics objects*, which are observed versions of real particles or groups of particles. Outstanding physics objects used in the analysis include leptons (electrons and muons), hadronic jets, and missing transverse energy, $\not\!\!E_T$. The physics objects are combined to reconstruct the $\ell^+\ell^-\mathrm{q\bar{q}}$ final state, allowing to study the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\,\mathrm{q\bar{q}}$ decay.

◇ ◇ ◇

The Higgs boson decay mode $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\,\mathrm{q\bar{q}}$ (ℓ=e, μ) presents a clear signature, two isolated same-flavour and opposite electric-charge leptons (electrons and muons) with invariant mass consistent with the Z boson mass, m_Z, and a pair of hadronic jets with invariant mass around m_Z. The absence of neutrinos and other potentially undetected particles allows a full kinematic reconstruction of the events. The mass of the $\ell^+\ell^-\mathrm{q\bar{q}}$ system, $m_{\ell\ell\mathrm{qq}}$, is a direct measurement of the Higgs boson mass, m_H, and is used as a statistical estimator to isolate a hypothetical signal. Background events with a topology similar to the Higgs boson decay are massively produced at the LHC, as they originate from SM processes with cross sections several orders of magnitude (up to 4) above the Higgs boson total production cross section. The main such processes are the production of a Z boson in association with jets (Z+jets), the production of top quark pairs ($\mathrm{t\bar{t}}$) and the diboson production (ZZ, WZ, and WW).

Efficient detection and identification of particles, as well as precise momentum and vertex determination, are key to a proper physical interpretation of the events.

Understanding the detector and accurately defining the physics objects used in the analysis are essential to avoid biases that either would render the signal undetectable or, even worse, would create fake signals.

3.1 Particle flow algorithm

The particle flow (PF) event reconstruction is an algorithm that aims at reconstructing and identifying individual particles in the event: electrons, photons, muons, charged hadrons and neutral hadrons [45]. This algorithm combines information from all CMS subdetectors to perform an optimal reconstruction of the physics objects, leading to an improved performance of the reconstruction of jets (section 3.3) and missing transverse energy (section 3.4), and of the lepton identification (section 3.2). Only particles reconstructed by the PF algorithm are used in the analysis.

The PF algorithm identifies photons as very narrow energy clusters in the ECAL, without an associated track measured in the silicon tracker. Electrons are detected as isolated charged tracks pointing to highly collimated clusters in the ECAL. Muons are identified as isolated tracks in the muon systems consistent with charged tracks reconstructed in the tracker, and associated to energy deposits in the calorimeters compatible with minimum ionization. Hadrons are made of the calorimetric energy clusters, both in the ECAL and the HCAL, that are not already paired to either photons, electrons or muons. Tracker information is used to distinguish charged from neutral hadrons.

The particles identified by the PF algorithm are used to build hadronic jets, to reconstruct the τ decays and to measure the missing transverse energy of the event, which is used to infer the presence of particles invisible to the CMS detector, such as neutrinos [46].

3.2 Lepton reconstruction

Muon tracks are reconstructed combining the information provided by the tracker system and muon stations. Only muons inside the acceptance volume of the tracker system, $|\eta| < 2.4$, are selected for the analysis. In order to ensure a precise momentum measurement, a set of standard quality criteria are applied to muons. They must be built from at least 5 hits on the silicon tracker, and at least two track segments (3-hit to 4-hit groups) reconstructed in the muon stations. The fit performed to extract the parameters of a helical track must have a χ^2 goodness-of-fit

normalized to the number of degrees of freedom less than 10. Good muon tracks are required to originate from the primary vertex, with transverse and longitudinal impact parameters $d_{xy} < 2$ mm and $d_z < 5$ mm, respectively. These criteria reject off-vertex muons, like secondary cosmic rays, in-flight meson decays and muons from hadronic interactions in the HCAL (punch-through muons).

Electrons are reconstructed using a Gaussian Sum Filter (GSF) algorithm [47] which combines a track in the tracker system and a supercluster in the ECAL consistent with an electromagnetic cascade. Only the electrons detected inside the ECAL acceptance volume, $|\eta| < 2.5$, are selected, excluding the gap region between the ECAL barrel and the endcaps ($1.44 < |\eta| < 1.57$) [37]. Two sets of quality criteria are applied to the electrons reconstructed, either in the barrel ($|\eta| < 1.48$) or in the endcap ($1.48 < |\eta| < 2.5$) (Table 3.1), based on the following variables: the angular separation between the ECAL supercluster and the track, $\Delta\phi_{trk}$ and $\Delta\eta_{trk}$; the shape of the supercluster in the η direction, $\sigma_{i\eta i\eta}$; the ratio between energy deposits in the ECAL and HCAL, H/E; the difference between the measured energy in the ECAL (E) and the momentum measured on the tracker (p), $|1/E - 1/p|$; the transverse and longitudinal distance to the primary vertex, d_{xy} and d_z and the isolation of the electrons from the hadronic activity of the event, I/p_T.

Table 3.1: Selection requirements for electrons.

Variable	Barrel region	Endcap region
$\Delta\phi_{trk}$	< 0.007	< 0.009
$\Delta\eta_{trk}$	< 0.15	< 0.1
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
H/E	< 0.12	< 0.10
$\lvert 1/E - 1/p\rvert$	< 0.05	< 0.05
d_{xy}	< 0.2 mm	< 0.2 mm
d_z	< 2 mm	< 2 mm
I/p_T	< 0.15	< 0.15

The hadronic production in proton-proton collisions is extremely high, including secondary decays of hadrons into muons and electrons. Hence, prompt electrons and muons from Z boson decays are required to be isolated with respect to the hadronic activity of the event. The lepton isolation (I) is defined as the sum of the transverse momenta of the charged hadrons (ch) and the transverse energy of the neutral hadrons (nh) and photons (γ), around a cone centred on the lepton direction with radius $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. It is convenient to define the isolation relative

to the total measured p_T of the lepton as:

$$\frac{I}{p_T} = \frac{1}{p_T}\left[\sum_{ch} p_T + \max\left(0, \sum_{nh} E_T + \sum_{\gamma} E_T - \mathrm{PU}\right)\right] . \tag{3.1}$$

The pile-up term (PU) accounts for the additional energy from proton-proton collisions other than the hard scattering. This extra term is extracted from the hadronic activity identified by the PF algorithm [48]. The cone radius used to estimate the isolation in electrons is $\Delta R = 0.3$ and the relative isolation is required to be I_e/p_T <15%. For muons, the cone radius is $\Delta R = 0.4$ and $I_\mu/p_T < 12\%$.

The lepton efficiency is one important aspect of the lepton identification and reconstruction and is quantified from observed data using a Tag and Probe method [49]. In this procedure, efficiencies are measured from a control sample of pure Z$\rightarrow \ell^+\ell^-$ events, constructed by imposing tight identification requirements on one of the leptons of the event, the so-called *tag*. The other leg is the *probe*, built pairing the remaining opposite-sign leptons in the event with the tag, fulfilling a loose lepton selection and requiring the dilepton mass to be in the interval [60,120] GeV. The efficiency is extracted as the number of probes passing the lepton identification criteria over the total number of probes in the sample. This method is also used to determine the efficiency of the isolation and trigger criteria.

Deviations of the efficiency estimated in the simulation (ϵ_{sim}) with respect to the efficiency measured in the observed data (ϵ_{data}) are corrected using multiplicative scale factors defined as:

$$\mathrm{SF} = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{sim}}} . \tag{3.2}$$

These scale factors are computed as functions of p_T and η for electrons and muons separately (Tables 3.2 and 3.3). The effect of this correction on the simulated samples is small, less than 2% in most of the phase space.

Table 3.2: Scale factors for the muon reconstruction efficiency.

Muon SF	0< $\lvert\eta\rvert$ < 0.9	0.9< $\lvert\eta\rvert$ < 1.2	1.2< $\lvert\eta\rvert$ < 2.1	2.1< $\lvert\eta\rvert$ < 2.4
0< $\lvert\eta\rvert$ < 0.9	0.987±0.003	0.987±0.003	1.007±0.003	1.023±0.006
0.9< $\lvert\eta\rvert$ < 1.2	0.987±0.003	0.997±0.012	1.008±0.005	1.007±0.014
1.2< $\lvert\eta\rvert$ < 2.1	1.007±0.003	1.008±0.005	1.014±0.009	1.010±0.011
2.1< $\lvert\eta\rvert$ < 2.4	1.023±0.006	1.007±0.014	1.010±0.011	1.036±0.027

Table 3.3: Scale factors for the electron reconstruction efficiency.

Electron SF	$0< \|\eta\| < 0.8$	$0.8< \|\eta\| < 1.44$	$1.56< \|\eta\| < 2.0$	$2.0< \|\eta\| < 2.5$
$p_T\in$ [20,30] GeV	0.982±0.002	0.993±0.002	0.988±0.005	1.002±0.004
$p_T\in$ [30,40] GeV	0.988±0.001	0.993±0.003	0.993±0.003	1.004±0.003
$p_T\in$ [40,50] GeV	0.990±0.001	0.993±0.002	0.992±0.002	1.005±0.002
$p_T>$ 50 GeV	0.990±0.001	0.986±0.002	0.990±0.003	0.998±0.004

3.3 Jet reconstruction

Due to colour charge antiscreening in QCD, free quarks and gluons produced in high energy collisions cannot exist individually. High energy densities achieved in quark and gluon steered colour strings give rise to quark-antiquark production from the vacuum, which condense into hadrons when the energy density drops below some scale, typically around 1 GeV. This process, known as hadronization, is responsible for the production of high multiplicity hadronic jets in high energy collisions, which are highly boosted in the LHC environment. In CMS, the pixel and silicon tracker, and the electromagnetic and hadron calorimeters, are key to detect these jets of particles, measure their energy and mass and provide information on the flavour of the primary quarks from which they originated.

Hadronic jets are reconstructed by clustering the particles detected in the calorimeters inside a cone of a given radius, R. Two complementary algorithms are used. The anti-k_t algorithm [50] uses R =0.5 (so-called AK5) and is optimized for the reconstruction of narrower jets from the hadronization of isolated high-p_T partons. The Cambridge-Aachen algorithm [51] with R =0.8 (named CA8) is used to identify highly boosted hadronically decaying Z bosons, which give rise to two overlapping jets that are reconstructed as single jets in the detector, known as merged jets. This algorithm provides information about the substructure of the jet, allowing to design sub-jet identification algorithms (details in the next section).

Jet quality criteria are applied to remove noise and fake jets, such as misidentified particles or pile-up jets: only jets inside the tracker acceptance volume $|\eta| < 2.4$ are accepted in the analysis; the fraction of energy in the HCAL or in the ECAL due to neutral hadrons must be less than 99%; the fraction of energy in the ECAL due to charged hadrons is required to be less than 99%; the jet has to contain at least two constituents, one of them charged; and part of the energy due to charged hadrons has to be detected in the HCAL.

To avoid misidentifying leptons as jets a certain angular separation, ΔR, is imposed on the reconstructed jets with respect to any isolated lepton in the event: $\Delta R > 0.5(0.8)$ for AK5 (CA8) jets.

Several effects can potentially bias the measured energy and resolution of the jets, such as the non-linear response of the calorimeter, soft particles entering in the jet cone or electronic noise in the detector. This bias is compensated by means of p_T and η-dependent jet energy corrections, which are estimated using γ+jet and Z+jet control data samples [52]. The uncertainty related to the jet energy scale is the dominant systematic uncertainty on the jet reconstruction, as discussed in section 4.7.

Jets identified in pile-up interactions are removed using information of their charged constituents. A pile-up mitigation variable, β, is defined as the sum of the transverse momenta of the charged particles in the jet coming from the primary vertex (ch=PV), divided by the transverse momenta of all charged particles in the jet (ch):

$$\beta = \frac{\sum_{ch=\mathrm{PV}} p_T}{\sum_{ch} p_T} \ . \tag{3.3}$$

Low values of β indicate that a significant amount of particles do not come from the hard interaction. The distribution of this variable for AK5 jets for the Z+jets background and for a Higgs boson signal with m_{H}=300 GeV is shown in Figure 3.1. Rejecting events with $\beta < 0.2$ mitigates the contribution of pile-up jets significantly.

3.3.1 Jet merging

The CA8 algorithm uses jet substructure methods to identify merged jets from boosted di-quarks as hadronic Z boson candidates and to reject pure QCD jets generated in single quark and gluon hadronization [53, 54].

The reconstructed mass of the jet is one of the key observables used to identify merged jets from Z bosons. A grooming technique, so-called *pruning* algorithm [55], reduces the impact of the underlying event, pile-up and low-p_T gluon contributions to the jet, giving a more accurate jet mass determination (Figure 3.2). The effect of the pruning is to shift the mass of QCD jets to lower values. At the same time, the reconstruction of boosted Z-boson jets is improved, showing better resolution and a smaller bias with respect to m_{Z}. The sub-jets identified by the CA8 algorithm inside the merged jets are used for the angular analysis of the events (section 4.5) and for the jet flavour identification (section 3.3.2), in a similar way to the analysis

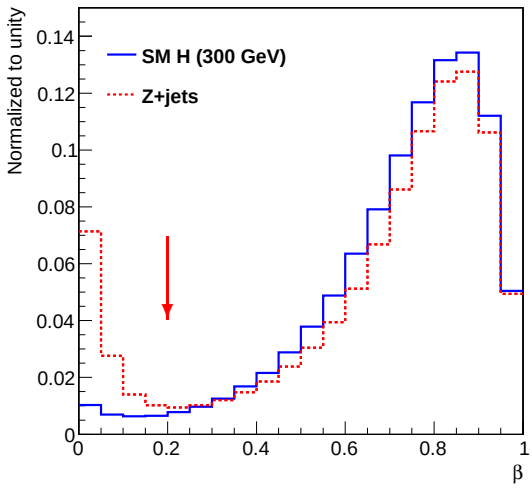


Figure 3.1: Pile-up mitigiation variable, β, for the AK5 jets of a Higgs boson signal of m_{H}=300 GeV (solid blue line) and for Z+jets events (dashed red line). The red arrow indicates the position of the selection requirement.

of events with resolved jets (AK5).

Merged jets reconstructed by the CA8 algorithm are subject to further quality requirements, in order to reject broad QCD jets not containing two distinct jets. The so-called *N-subjettiness* variable [56], τ_{N}, is related to the probability for a jet to be constituted by N sub-jets:

$$\tau_{\mathrm{N}} = \frac{1}{\sum_i p_{T,i} R_0^{\delta}} \sum_i p_{T,i} \cdot \min\{(\Delta R_{1,i})^{\delta}, (\Delta R_{2,i})^{\delta}, ..., (\Delta R_{\mathrm{N},i})^{\delta}\} \ , \tag{3.4}$$

where i runs over all the jet constituents and $\Delta R_{n,i}$ is the angular distance of the i^{th} constituent to the axis of the n^{th} sub-jet. The precise values of the parameters, $R_0 = 0.8$ and $\delta = 1$, are given by the CA8 clustering algorithm. Values of τ_N close to zero correspond to a merged jet consistent with the N-sub-jet hypothesis.

In the analysis, the requirement $\tau_{21} \equiv \tau_2/\tau_1 < 0.5$ is used to isolate genuine boosted Z boson decays, with two sub-jets, from single-jet QCD events (Figure 3.3). This condition enhances the sensitivity to high mass Higgs boson signals.

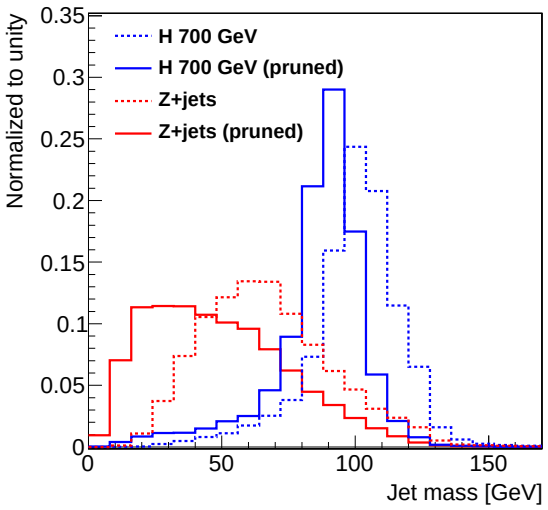


Figure 3.2: Invariant mass distribution of CA8 jets before (dashed line) and after (solid line) pruning. The blue histograms correspond to a Higgs boson signal of 700 GeV mass, while the red histograms are Z+jets events.

3.3.2 Jet flavour identification

The presence of hadronically decaying Z bosons is an important difference between the Higgs boson signal and the Z+jets events (where Z$\rightarrow\ell^+\ell^-$). The jet flavour population of the Z+jets background is dominated by jets originating in the hadronization of gluon or proton valence quarks (u, d). The Z bosons decay in approximately equal fractions of u, d, s, c and b quarks. The much higer relative fraction of b quarks in the Higgs boson events with respect to the background permits to enhance the signal content using jet-flavour tagging techniques. (Figure 3.4).

Several algorithms, generically known as *b-tagging* [34], exploit the relatively long lifetime of B hadrons to identify jets coming from the hadronization of b quarks, referred to as *b jets*. The mean path traveled by B hadrons is long enough, up to few mm, to make feasible to reconstruct secondary decay vertices inside the hadronic jets, displaced with respect to the primary vertex of the collision. This is possible thanks to the excellent precision of the tracker system. Jets containing bottom quarks are also characterized by larger impact parameters of the tracks, higher charged-track multiplicities and larger jet masses with respect to the light-quark jets.

The Jet Probability algorithm [57] calculates, for each track i in the jet, its probability to originate from the primary vertex, P_i, based on the impact parameter of

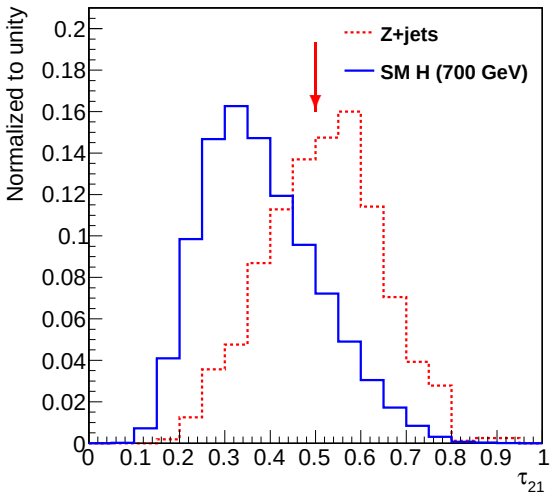


Figure 3.3: N-subjettiness ratio, τ_{21}, for a 700 GeV mass Higgs boson signal (solid blue line) and for Z+jets events (dashed red line). Boosted hadronic Z bosons are more likely in signal events. The red arrow indicates the position of the selection requirement.

the track. A likelihood function is defined for each jet as:

$$\mathcal{L}_j = \prod_i P_i \cdot \sum_i \frac{(-ln\,[\prod_i P_i])^i}{i!}\ . \tag{3.5}$$

The negative logarithm of this function, JP, is used as a b-tagging discriminator:

$$\text{JP} = -ln\ \mathcal{L}_j \tag{3.6}$$

High values of the discriminator are more likely for jets containing a B hadron, while low values are associated with light quarks and gluons, as displayed in Figure 3.5.

The nature of the flavour-tagging algorithms makes them penalize the purity of the b jet selected samples when a high identification efficiency is pursued. Conversely, a high purity sample suffers from a lower efficiency. The discriminator value used to tag b jets can be chosen to maximize one or the other, leaving the analyser the freedom to define samples with different sensitivities to a potential signal. A CMS standard requirement of JP>0.275 (called *loose working point*, or JPL) retains more b jets (85% efficiency), with a higher rate of light-quark jets misidentified as b jets, around 10%. A pure b jet sample, with misidentification of light jets of 1%, is selected with JP>0.545 (*medium working point*, or JPM), with a lower efficiency though, around 60%.

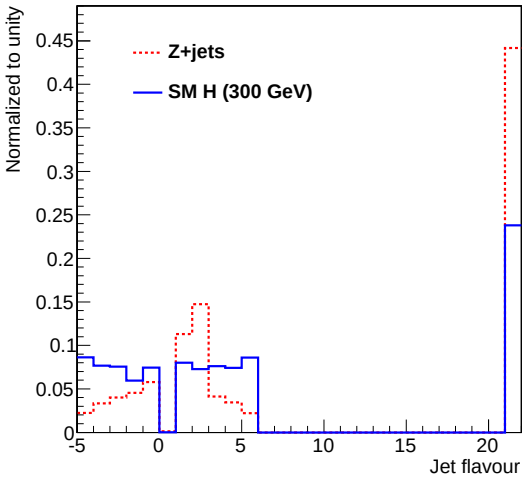


Figure 3.4: Jet flavour code for simulated Z+jets events and for a 300 GeV mass Higgs boson signal. The numbers 1, 2, 3, 4, 5 denote d, u, s, c, b quarks, antiquarks are negative and gluon is 21.

The sensitivity of the analysis to a Higgs boson signal is enhanced classifying the events into three exclusive categories according to the number of b-tagged jets (chapter 4). The most sensitive b-tagged category contains events with a (sub-)jet passing the JPM working point and another (sub-)jet fullfiling the JPL working point criterion. This so-called 2-btag category has a low number of events, but a good signal-to-background ratio. The rest of the sample is classified in two more categories: the 1-btag category, where at least one of the (sub-)jets verifies the JPL working point requirement, and the 0-btag category, which includes the remaining events. The last category contains a large number of events, but its signal-to-background ratio is poorer.

In general, the reconstruction of simulated data tends to identify b jets more efficiently ($\epsilon_{\mathrm{b}}^{\mathrm{sim}}$) than in observed data ($\epsilon_{\mathrm{b}}^{\mathrm{data}}$). Therefore, multiplicative scale factors are used to correct the differences in efficiency. Two kinds of scale factors are applied [58], one for correcting the b jet identification efficiency,

$$\mathrm{SF}_{\mathrm{b}} = \epsilon_{\mathrm{b}}^{\mathrm{data}} / \epsilon_{\mathrm{b}}^{\mathrm{sim}} \ , \tag{3.7}$$

and another one for rectifying the mis-identification of light jets as b jets:

$$\mathrm{SF}_{\mathrm{mis}} = \epsilon_{\mathrm{mis}}^{\mathrm{data}} / \epsilon_{\mathrm{mis}}^{\mathrm{sim}} \ . \tag{3.8}$$

These scale factors are computed [57] as a function of the p_T and η of the jets and are

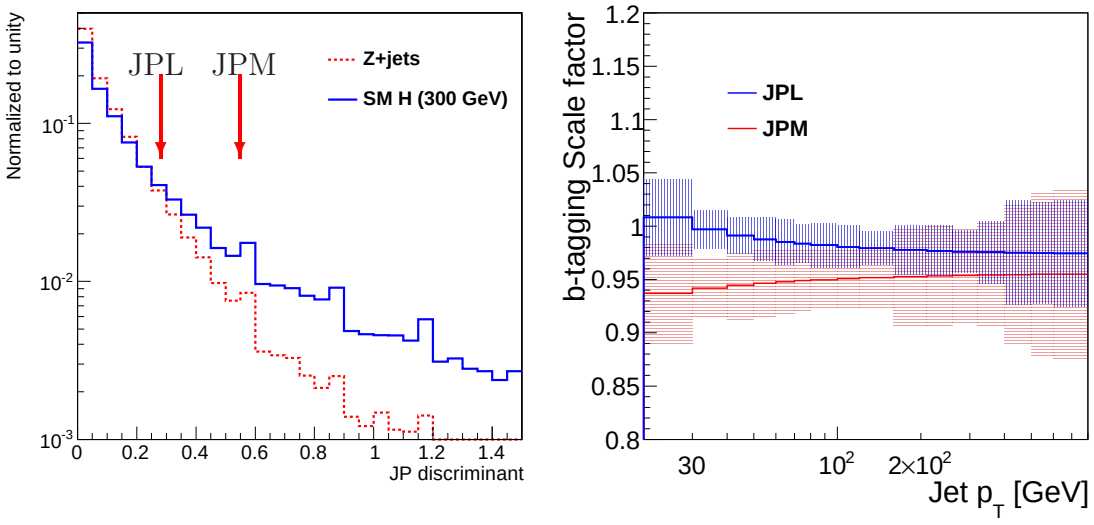


Figure 3.5: Left: JP discriminator for reconstructed AK5 jets in a 300 GeV mass Higgs boson signal (solid blue line) and for Z+jets events (dashed red line). The red arrows point to the position of the different working point requirements, JPL and JPM. Right: Scale factors for the two b-tagging working points, JPM (red line) and JPL (blue line) and their associated uncertainties (coloured bands around the lines).

found to be close to 1 (Figure 3.5, right), indicating that the b-tagging efficiencies are well reproduced in the simulation.

3.4 Missing transverse energy

Events with neutrinos in the final state, like $t\bar{t}$ production, are characterized by large missing tranverse energy values, $\not\!\!E_T$, the momentum imbalance in the transverse plane. The $\not\!\!E_T$ is defined as the vectorial sum of the transverse momenta of all the particles detected in the event [59]:

$$\not\!\!E_T = -\sum \vec{p_T} \ . \tag{3.9}$$

Events without neutrinos also present some amount of $\not\!\!E_T$ due to resolution or instrumental effects, such as the finite resolution of the calorimeters or regions not covered by the detectors, like cracks present in certain points to accomodate the electronics, cooling pipes and so on. The $\not\!\!E_T$ significance variable, λ, is introduced to quantify the likelihood that the observed $\not\!\!E_T$ in the event is consistent with a fluctuation from zero due to detector-related limitations [60]:

$$\lambda = -2 \cdot ln \frac{P(\not{E}_T = \not{E}_T^{\text{observed}})}{P(\not{E}_T = 0)} \ , \tag{3.10}$$

where $P(\not{E}_T = \not{E}_T^{\text{observed}})$ is the probability that the observed $\not{E}_T$ is due to undetected neutrinos and $P(\not{E}_T = 0)$ is the probability that the observed $\not{E}_T$ is an experimental effect. Large values of λ indicate a large probability of real $\not{E}_T$, while low values of the significance are related to instrumental and resolution effects. Figure 3.6 compares the λ distribution for $t\bar{t}$ events and a Higgs boson signal.

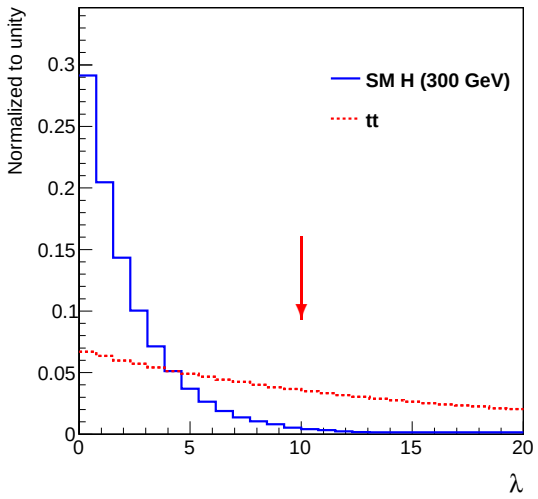


Figure 3.6: Distribution of λ for a Higgs boson signal of m_{H}=300 GeV (solid blue line) and $t\bar{t}$ events (dashed red line). The red arrow indicates the position of the selection requirement.

3.5 Reconstruction of Z and Higgs bosons

Pairs of same-flavour opposite-charge leptons reconstructed in the event are treated as $Z \to \ell^+\ell^-$ candidates. To ensure high lepton identification efficiency and good momentum resolution, sufficiently high momentum leptons are required. The transverse momentum of the hardest lepton must be above 40 GeV (leading lepton), while the other has to be at least 20 GeV (subleading lepton).

Leptonic Z boson candidates are accepted in the analysis if the reconstructed dilepton invariant mass, $m_{\ell\ell}$, lies inside a narrow window around the Z boson mass,

$|m_{\ell\ell} - m_{\mathrm{Z}}| < 15$ GeV. This requirement allows to largely reduce non-resonant backgrounds, such as the $\mathrm{t\bar{t}}$ production (Figure 3.7).

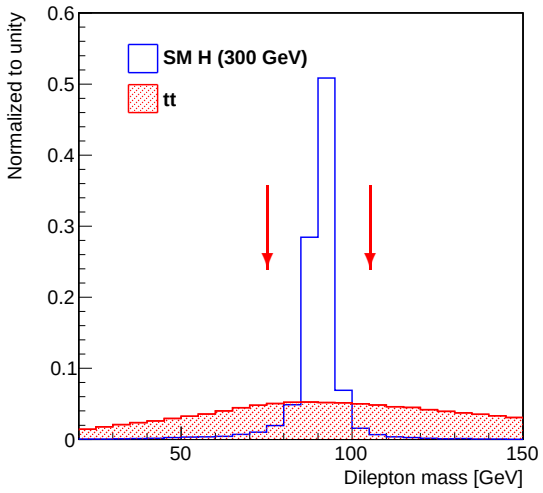


Figure 3.7: Dilepton invariant mass for a Higgs boson signal of m_{H}=300 GeV (blue histogram) and $\mathrm{t\bar{t}}$ events (red hatched histogram). The red arrows indicate the selected region.

Hadronic Z boson candidates, $\mathrm{Z}\rightarrow\mathrm{q\bar{q}}$, are reconstructed from pairs of AK5 jets. All the jets are required to have a p_T above 30 GeV, ensuring a precise momentum determination. The experimental resolution on the dijet pairs mass, m_{jj}, is worse than in the dilepton case, 11% versus 2%, and thus the events are selected in a broader m_{jj} window around the Z boson mass, $|m_{jj} - m_{\mathrm{Z}}| < 20$ GeV. These requirements on the jets p_T and m_{jj} largely reduce the background from jet pairs not coming from a Z boson decay. The dijet invariant mass distributions for the Z+jets events and for a 300 GeV mass Higgs boson signal are displayed in Figure 3.8.

The side bands of the m_{jj} distribution, adjacent to the signal region, are used as a signal-depleated control region. Less than 10% of the expected signal events lies in the control region, defined as 60 GeV $< m_{jj} <$ 71 GeV and 111 GeV $< m_{jj} <$ 130 GeV. More than 95% of the events in this region are Z+jets events. The control region is constructed in such a way that the kinematic properties of the background events are equivalent in the signal and control regions.

All the leptonic and hadronic decays of Z bosons are used to build Higgs boson candidates (named $\ell^+\ell^-\mathrm{q\bar{q}}$ candidates). The mass resolution of these candidates is

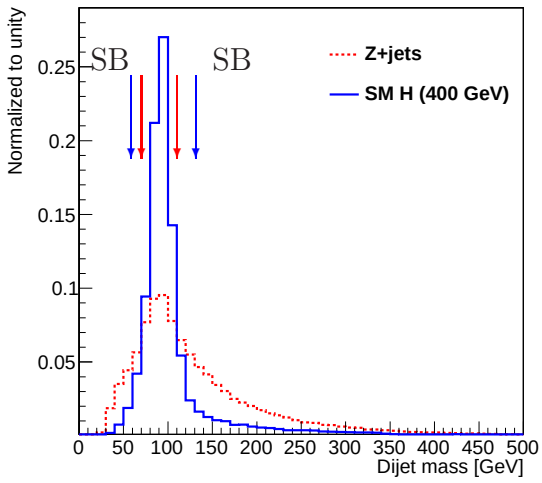


Figure 3.8: Dijet mass distribution for a 400 GeV SM Higgs boson (solid blue line) and for the main background, Z+jets (dashed red line). The central region 91±20 GeV, between the red arrows, is used to extract the hypothetical signal, while the adjacent side bands excluding the signal region, between the blue arrows, are the control region.

largely dominated by the dijet mass resolution. The goodness of the reconstruction of the Higgs boson is verified using simulated samples, which evidence a small bias, around 1-2 GeV, with respect to the nominal m_{H} value and a resolution around 5%. As an example, Figure 3.9 depicts the comparison between the simulated and the reconstructed mass for a Higgs boson signal of m_{H}=500 GeV.

In the case of reconstructed merged jets, each pair of leptonic and boosted hadronic Z bosons are used to build Higgs candidates (called $\ell^{+}\ell^{-}$J candidates), where J represents the CA8 merged jet formed by two sub-jets. In this subsample, the signal region is defined requiring the mass of the pruned merged jet to be close to m_{Z}, $m_J \in [71, 111]$ GeV. The control region, $m_J \in [60, 130] \cup [111, 130]$ GeV, which excludes the m_{Z} region, is used for background validation and normalization.

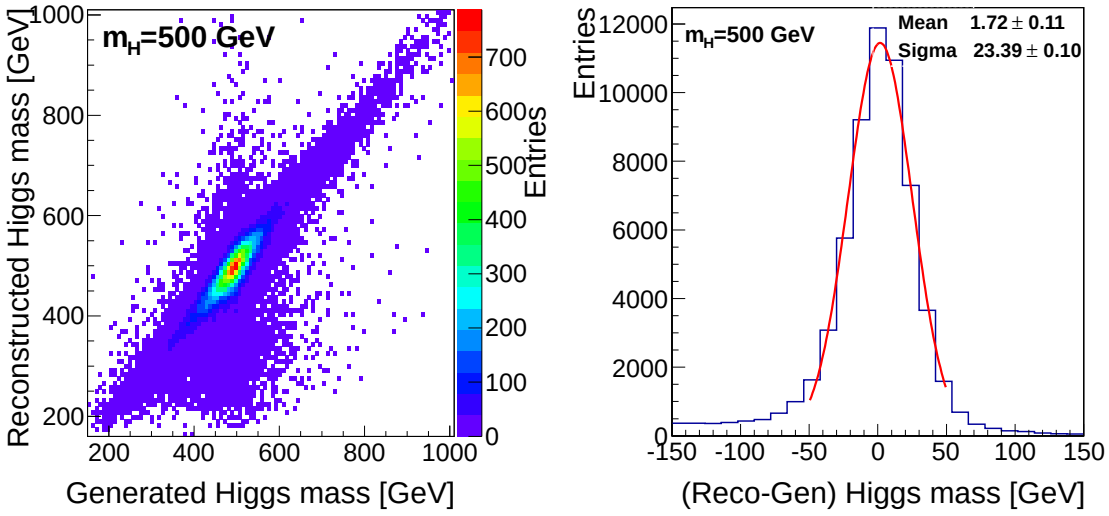


Figure 3.9: (Left) Comparison between generated and reconstructed invariant mass for a SM Higgs boson of $m_{\text{H}} = 500$ GeV. (Right) Reconstructed minus generated invariant masses for the same H mass hypothesis. The line represents a gaussian fit to the core of the distribution.

◇ ◇ ◇

Leptons and hadronic jets are efficiently detected and reconstructed in events collected by CMS to build Z→q$\bar{\text{q}}$, Z→$\ell^+\ell^-$and H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ candidates. The next chapter describes in detail the analysis performed in order to isolate hypothetical Higgs boson signals over the expected SM backgrounds, exploiting the kinematic properties of the H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ events and the different topologies (b-tagging, resolved and merged jets, production mechanism) to achieve the highest signal sensitivity.

Chapter 4

Analysis of the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\bar{q}}$ decay channel

Finding minute signals of new physics in the large dataset delivered by the LHC is a task of the utmost complexity. Profound knowledge of the underlying physics, both in the context of the SM and theories beyond, allows to design sophisticated analysis techniques that exploit the subtleties of the events kinematics to enhance a hypothetical signal to the detriment of the SM background. To put it in simple terms, the goal of the analysis is to increase the ratio of signal to background events in selected regions of the phase space. It is equally important to construct a robust statistical method capable of quantifying deviations of the data from the reference models. This chapter presents a detailed analysis of the search for Higgs bosons in the decay channel $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\,\mathrm{q\bar{q}}$ in CMS.

4.1 Overview of the analysis

The events selected to study the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\,\mathrm{q\bar{q}}$ decay channel present two isolated opposite-charged same-flavour leptons (ℓ=e, μ) with invariant mass consistent with m_{Z} and two hadronic jets with invariant mass also consistent with m_{Z} (Figure 4.1).

In order to isolate a hypothetical Higgs boson signal, the mass of the $\ell^+\ell^-\mathrm{q\bar{q}}$ system, $m_{\ell\ell\mathrm{q\bar{q}}}$, is used as discriminating observable (Figure 4.2). Above 200 GeV, this distribution shows an exponentially falling behaviour for the background, while the H resonance is distributed around the mass value of the signal hypothesis with a mass-dependent width (Figure 1.3 in section 1.2). This analysis searches for Higgs bosons in the mass range from 200 GeV to 1 TeV. The choice of the lower value is driven by the drop of the trigger efficiency for low transverse momentum objects. On the other side, the 1 TeV upper value is an a priori choice considered appropriate for the electroweak symmetry breaking scale, besides other theoretical limitations

on the calculation of the Higgs boson width [61, 62].

For high Higgs boson masses (m_{H}>600 GeV), quarks stemming from a hadronically decaying Z boson are strongly boosted, giving rise to overlapping jets which are reconstructed as a merged CA8 jet (Figure 4.3, upper).

The signatures described above are expected for signal events in which the Higgs boson is produced via the the gluon fusion mechanism (ggH) through a third generation quark loop (Figure 1.2 in section 1.2 and events in Figure 4.1). The sub-dominant production mode in proton-proton collisions is the vector boson fusion mechanism (VBF), in which the Higgs boson is produced in association with two relatively low transverse momenta jets stemming from the light quarks that participate in the hard interaction. Due to the low momentum transfer in this t-channel process, the two additional jets are produced in opposite directions along the z-axis, very close to the proton beam pipe (Figure 4.3, lower). This characteristic signature provides means of differentiating ggH and VBF events, leading to an increase in the sensitivity to a hypothetical signal, and permitting to study models where one of the production modes is either enhanced or suppressed.

A higher sensitivity to Higgs boson signals is achieved by splitting the data sample into exclusive categories, constructed from events with different physics objects and production mechanisms: dielectrons and dimuons, dijets and merged jets, b-quark jets and light-quarks jets, and eventually the presence of two additional high-η jets that differentiate between ggH and VBF production modes. The statistical data analysis (chapter 5) exploits these distinct topologies, accounting for their different systematic uncertainties and correlations. The observed data and the expected background are interpreted under the scope of different models, whether the SM or beyond, which yield unlike signal predictions, both in size and shape. A robust statistical method quantifies deviations of the data from the predictions allowing either to claim a discovery or, in the absence of significant deviations, to set exclusion limits on the production cross section and other model parameters.

The ultimate reach in sensitivity is accomplished by combining all the high mass Higgs boson searches performed by CMS with observed data in the decay modes H→ZZ and H→WW [63].

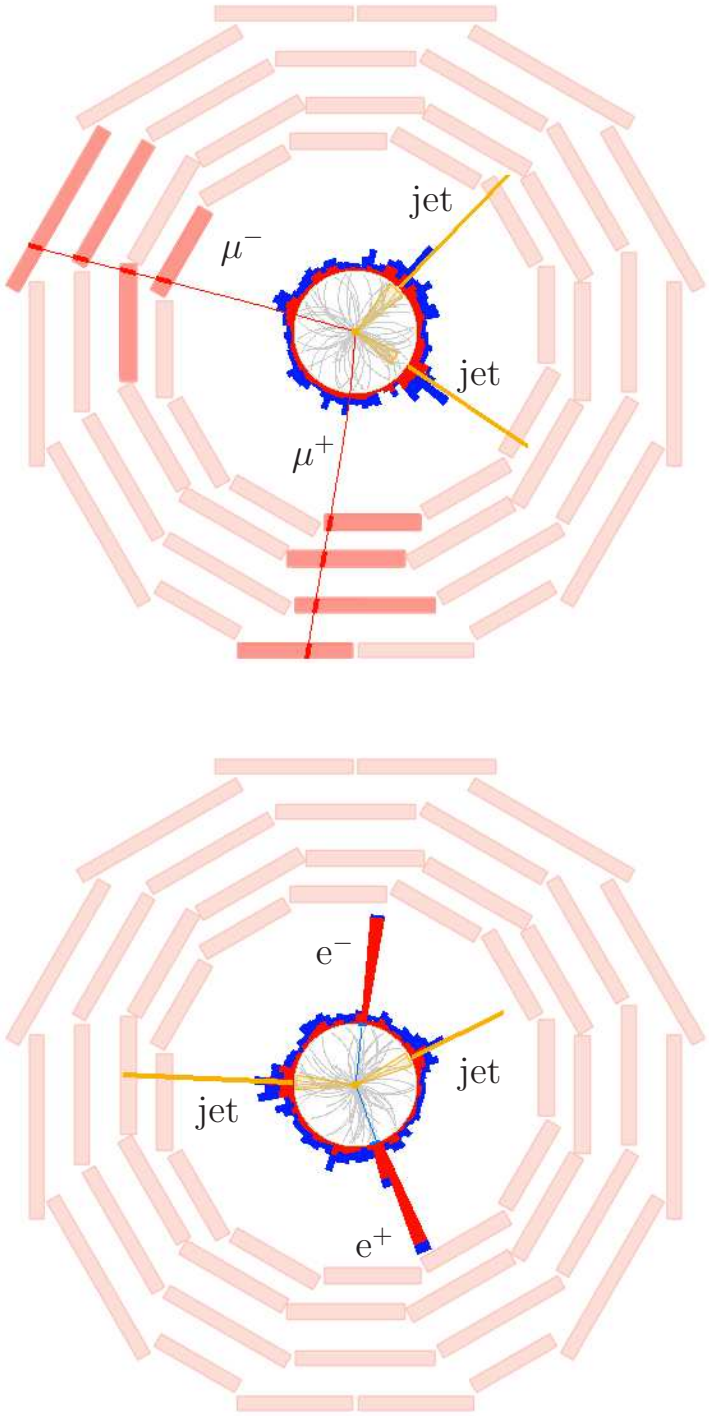


Figure 4.1: H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ candidates recorded by CMS. Upper: event with two hadronic jets detected in the calorimeters (yellow cones) and two muons passing through the DT chambers (red lines). Lower: event with two electrons detected in the ECAL (narrow blue spikes ending in the red towers) and two hadronic jets.

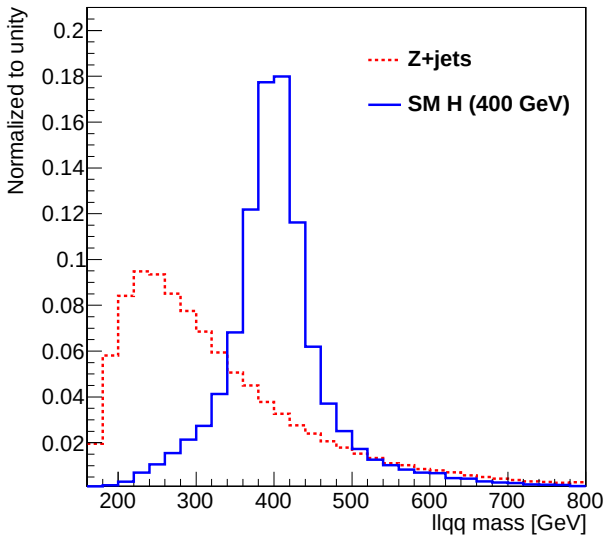


Figure 4.2: Invariant mass distribution of the $\ell^+\ell^-\mathrm{q\bar{q}}$ system for the main background, Z+jets (dashed red line), and a SM Higgs boson signal of m_{H}=400 GeV (solid blue line).

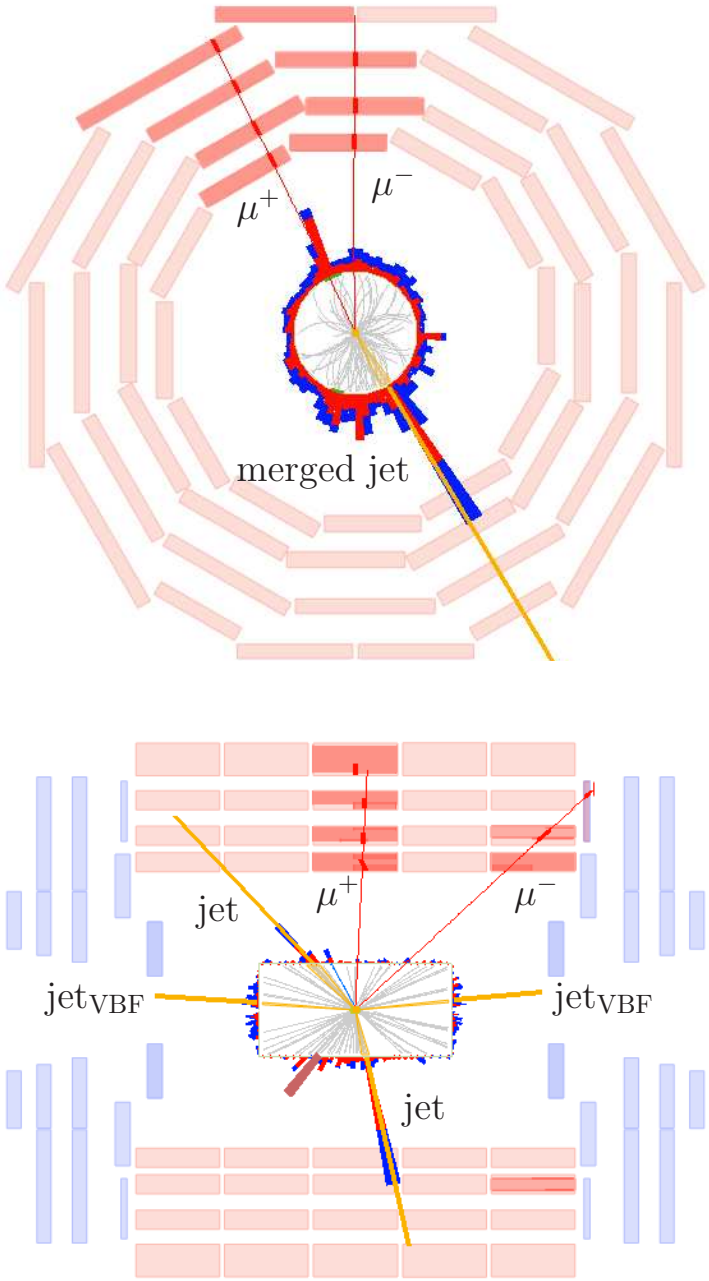


Figure 4.3: H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ candidates recorded by CMS. Upper: event with two muons and a high-p_T jet identified as a boosted Z boson candidate. Lower: event with two muons and four jets, two of them in the forward direction, identified as a VBF event.

4.2 Data sample

This analysis uses the data collected by CMS from proton-proton collisions produced in the LHC during the 2012 data taking period, at a centre-of-mass energy of $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 19.7 fb^{-1} [64].

Events passing the di-muon and di-electron trigger requirements are selected for their posterior analysis, that is, events with either two muons or two electrons satisfying p_T^1>17 GeV and p_T^2>8 GeV, where 1 and 2 indicate transverse momentum ordering. Furthermore, events with a muon and an electron with p_T>17 GeV and p_T>8 GeV, the e-μ trigger, are selected for background studies (section 4.6). The offline requirements of the analysis are tighter than the trigger conditions.

The number of interactions per event (pile-up) in the 8 TeV CMS data is estimated from the instantaneous luminosity delivered by the LHC (Figure 4.4). The simulations described in section 4.3 are generated prior to the data taking, hence before the actual pile-up distribution is known. Therefore, the events simulated are generated with an estimated pile-up distribution, and corrected a posteriori using an event-by-event weight to match the distribution of the number of proton-proton interactions in data, determined using the bunch-to-bunch instantaneous luminosity measurements and the proton-proton inelastic cross section. To test the result of the weighting procedure, the distribution of reconstructed vertices in observed data and in simulated events is compared (Figure 4.4, right) and an excellent agreement is observed.

4.3 Background simulation

A precise understanding of the overwhelming contribution of the different backgrounds is a crucial element of the analysis. The kinematic properties and yields of the SM processes that constitute the background events of this analysis, as well as the predicted heavy Higgs boson signals, are extracted from precise calculations performed at next-to-leading order (NLO) in perturbation theory, or even at higher orders (NNLO, NNLL, N3LO) depending on the physical process. The practical impossibility to model analytically the effects of the trigger and selection requirements used to isolate the signals, as well as the detector resolution and inefficiencies, forces the analysis to resort to simulated data produced using Monte Carlo techniques for the physics studies. These simulated data are compared to the observed data in order to test for the presence of signals over the SM background expectation, consistent with the properties of the Higgs boson. To reduce the statistical uncertainty of the predictions, large numbers of simulated events are generated, and later scaled

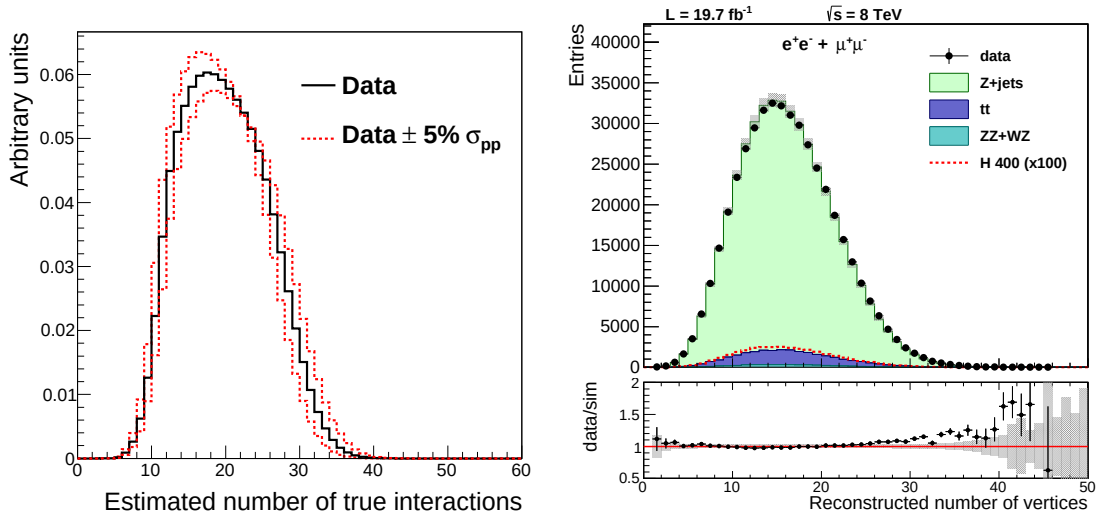


Figure 4.4: Left: estimated number of proton-proton interactions for $\sqrt{s} = 8$ TeV CMS data. The dashed red lines correspond to the up and down uncertainties, computed varying the proton-proton cross section by ±5%. Right: reconstructed number of primary vertices for observed data (dots with error bars) and for the background (filled histograms). The background is normalized to the data. The lower panel presents the ratio of the observed data to the simulation, along with the systematic uncertainties..

to the cross section of each process.

The main background to the Higgs boson signal is the production of a Z boson in association with jets. Several samples of Z+jets events are simulated using the MadGraph program at NLO [65] in order to enhance the numbers of events in certain regions of the phase-space, allowing to perform reliable comparisons with high statistical significance. The numbers of simulated events are between 3 and 20 times larger than the number of Z+jets events expected. The specific samples include high-multiplicity events (with up to 4 partons in the final state), boosted events with high-p_T leptons stemming from a Z boson, and events with high H_T, where $H_T = \sum_{\text{jets}} p_T$.

Simulated jets are constructed from stable hadrons resulting from the hadronization of the final state partons. This hadronization is modelled numerically with the program Pythia6 at NLO [66], which incorporates a variety of algorithms related to parton showering and matrix element calculations, and tunable parameters that allow to reproduce QCD-related phenomenology in a way consistent with the observed data in jet-enriched samples. The parton distribution functions (PDF) are

modeled in the simulations using two different parametrizations: CT10 at NLO [67] and CTEQ6 at LO [68].

Another important background contribution comes from $t\bar{t}$ production, where the leptonic decay of the top yields two opposite-sign leptons and two b jets in the final state. The simulation of the $t\bar{t}$ process is generated with the POWHEG program at NLO [69, 70]. The number of simulated $t\bar{t}$ events is 1.5 times larger than the yield expected in the collected data.

Diboson production (ZZ, WZ and WW) represents a minor contribution of the expected background. However, their cross sections are more than one order of magnitude higher than the predicted yields for the signals considered in the analysis. Ten times more events than expected are simulated with the Pythia program.

The list of background samples and their cross sections is given in Table 4.1.

Table 4.1: Summary of simulated background samples. The first column indicates the process produced using the program indicated in the last column. The second column lists the selection requirements used at generator (parton/particle) level when producing the samples. The third column shows the cross section of each process in pb.

Process	Generator cuts	σ (pb)	Program
Inclusive Z+jets	M($\ell^+\ell^-$) > 50 GeV	3503.7	MadGraph
Exclusive Z+1jet	M($\ell^+\ell^-$) > 50 GeV & N(partons)=1	660.6	MadGraph
Exclusive Z+2jets	M($\ell^+\ell^-$) > 50 GeV & N(partons)=2	215.1	MadGraph
Exclusive Z+3jets	M($\ell^+\ell^-$) > 50 GeV & N(partons)=3	65.8	MadGraph
Exclusive Z+4jets	M($\ell^+\ell^-$) > 50 GeV & N(partons)=4	27.6	MadGraph
High H_T Z+jets	200< H_T <400 GeV	23.4	MadGraph
Very high H_T Z+jets	H_T >400 GeV	3.4	MadGraph
High p_T Z+jets	p_T($\ell^+\ell^-$) >100 GeV	39.1	MadGraph
Very high p_T Z+jets	p_T($\ell^+\ell^-$) >180 GeV	5.5	MadGraph
$t\bar{t}$	–	245.8	POWHEG
ZZ	–	17.7	Pythia
WZ	–	22.9	Pythia
WW	–	57.1	Pythia

4.4 Signal modelling

Large samples of Higgs boson signal events are simulated at NNLO with the POWHEG program [70] for various Higgs boson mass hypotheses, m_{H}, in the range from 200 GeV to 1 TeV. The simulated decay, H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$, includes all lepton flavours (e, μ and τ). Exclusive samples are produced for the two production mechanisms ggH and VBF. Additionally, samples of Higgs bosons produced by the ggH mechanism in association with two partons are generated using the MINLO program [71], which incorporates a NLO calculation. These large samples are suitable for studies of events with additional jets (detailed in section 4.5). The cross section times decay branching fraction of the different simulated Higgs boson samples with distinct m_{H} values is shown in Figure 4.5.

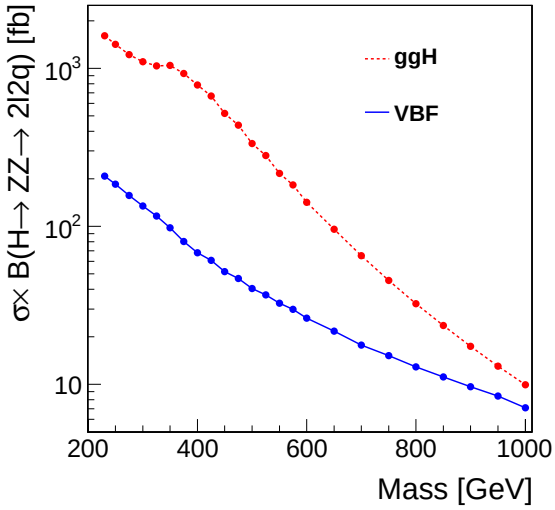


Figure 4.5: Cross section times branching fraction, $\sigma \times B$(H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$), as a function of the Higgs boson mass, both for the ggH (dashed red line) and VBF (solid blue line) production mechanisms. The dots are drawn at the precise mass values used in the simulated signal samples.

The POWHEG computation uses the so-called zero-width approximation, where the lineshape is assumed to be a Breit-Wigner distribution of width Γ (Figure 4.6, left). The uncertainty on the zero-width approximation is of the order of Γ/m_{H} [72], being only adequate for relatively low Higgs boson masses, where the width is expected to be small compared to its mass. For m_{H} above 400 GeV, the simulated mass is corrected using the complex-pole scheme (CPS) approximation [73], affecting both the width and the mean mass value of the Higgs boson lineshape (Figure 4.6, right).

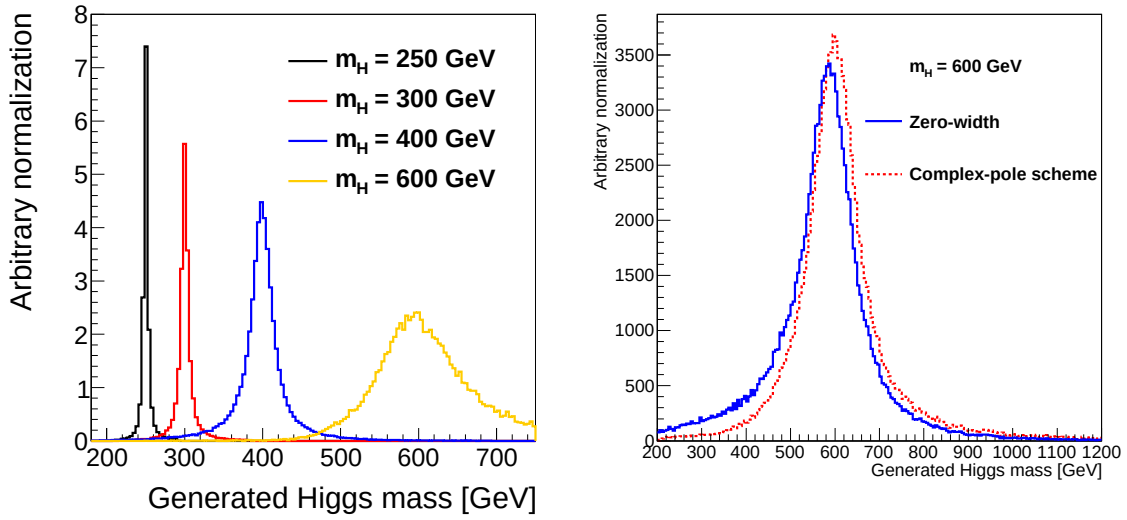


Figure 4.6: Left: Simulated mass distributions for several m_{H} values using the zero-width approximation: 250 GeV (black), 300 GeV (red), 400 GeV (blue) and 600 GeV (orange). Right: Different lineshapes for the zero-width (solid blue line) and CPS (dashed red line) approximations for a Higgs boson of m_{H}=600 GeV.

Furthermore, there exists a SM process with gluons in the initial state producing two Z bosons, the non-resonant gg→ZZ production, mediated by a box diagram at lowest order (Figure 4.7, left). This process interferes with the Higgs boson produced by the gluon fusion mechanism, specially for masses larger than 400 GeV [74]. The interference affects the lineshape constructively below the value of the mass peak and destructively above. The change in the Higgs boson invariant mass with respect to the CPS lineshape is shown in Figure 4.7, right.

4.5 Event selection

In order to search for heavy Higgs bosons, $\mathrm{H} \rightarrow \mathrm{ZZ} \rightarrow \ell^+\ell^- \mathrm{q\bar{q}}$ events are selected based on the reconstructed objects described in chapter 3. The purpose of the selection procedure is to increase the sensitivity to a hypothetical Higgs boson signal, which is achieved by reducing the background contamination using topological and kinematic information of the event, keeping as many of the signal events as possible.

Higgs boson candidates are built from all possible $\ell^+\ell^-$jj and $\ell^+\ell^-$J combinations in the event. The lepton and jet selections are detailed in chapter 3. To select events

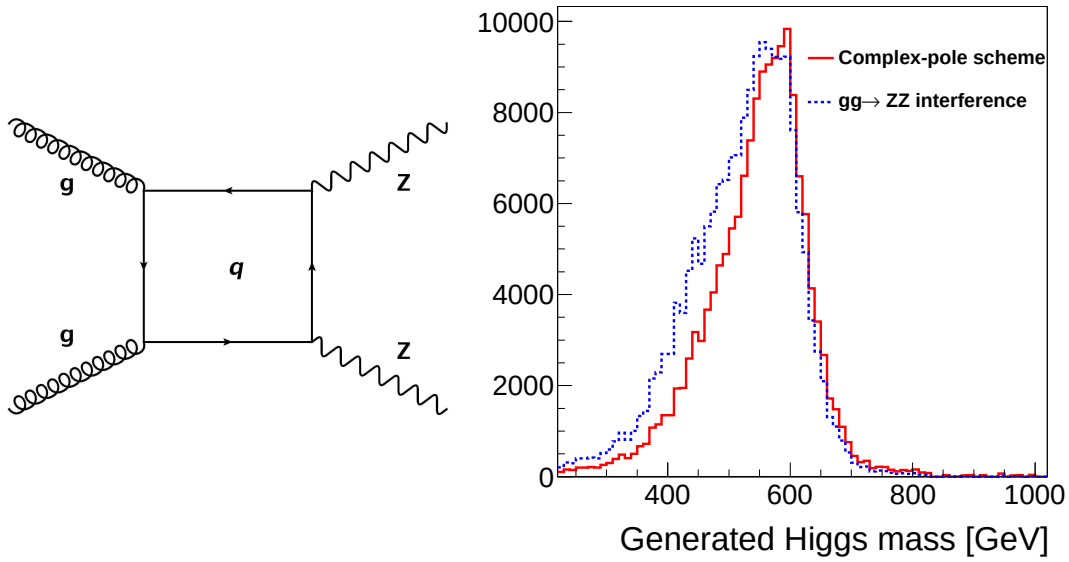


Figure 4.7: Left: Box diagram corresponding to gg→ZZ production. Right: Higgs boson lineshapes for the CPS correction (solid red line) and taking into account the gg→ZZ interference (dashed blue line) for a 600 GeV Higgs boson signal.

with a leptonic Z boson (section 3.5) the dilepton invariant mass $m_{\ell\ell}$ is restricted to be within a mass window around the Z boson mass, $|m_{\ell\ell} - m_{\text{Z}}| < 15$ GeV. The dilepton invariant mass depicted in Figure 4.8, left, displays a fair agreement between the observed data and the expected background.

To select events with a hadronic Z boson decaying to a pair of resolved jets ($\ell^+\ell^-$jj candidates), the dijet invariant mass is required to lie in an interval around the Z boson mass, $|m_{jj} - m_{\text{Z}}| < 20$ GeV, where most of the signal events are concentrated. The dijet invariant mass distribution (Figure 4.8, right) observed in data displays an excellent agreement with the background expectation.

In order to suppress backgrounds with neutrinos in the final state (section 3.4), events are required to have $\lambda < 10$. The λ distribution for observed data and the expected backgrounds is shown in Figure 4.9. This selection reduces the Higgs boson signal only by 1%, while the backgrouns with genuine $\not\!\!E_T$ are greatly reduced. Although the shape of the background is well reproduced by the simulation for events without true $\not\!\!E_T$, a scaling of 15% is applied to λ to correct for imperfect modelling of the jet resolution. The effect of this correction is treated as a systematic uncertainty on the statistical analysis.

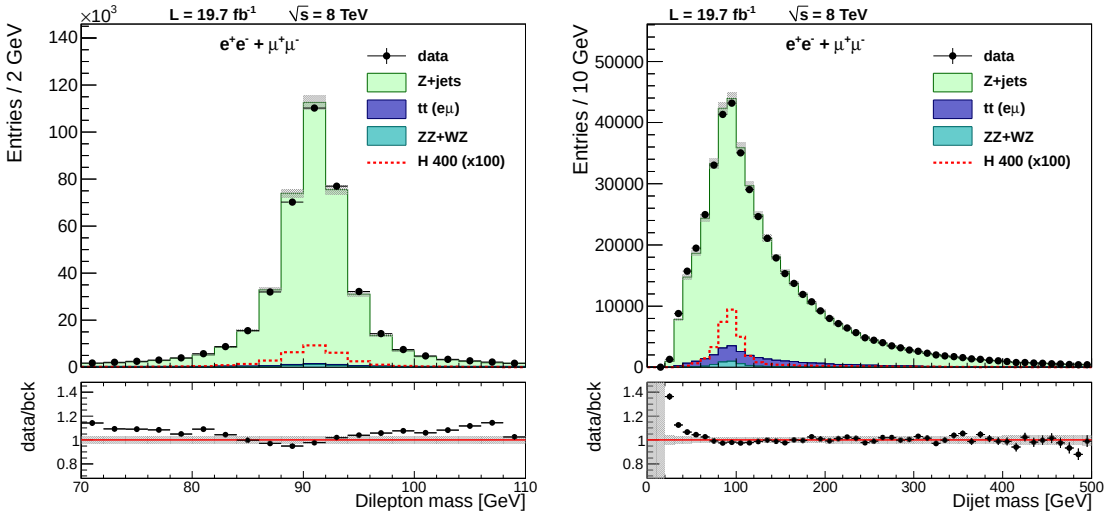


Figure 4.8: Distribution of the dilepton invariant mass (left) and the dijet invariant mass (right) for $\ell^+\ell^-\mathrm{q\bar{q}}$ events selected before the event selection. Dots with error bars correspond to the observed data, while the background is represented as filled histograms. In addition, the red histogram depicts the distribution of a Higgs boson signal (m_{H}=400 GeV), scaled by a factor 100 for better visibility. The lower panel shows the ratio between the observed data and the expected background, together with the systematic uncertainties.

Higgs boson candidates with a boosted hadronic Z boson ($\ell^+\ell^-$J) are built from dilepton pairs and CA8 merged jets. The lepton selection is the same as in the $\ell^+\ell^-$jj analysis. As discussed in section 3.3.1, for this kind of events, the Z bosons are significantly boosted. Thus, only events with $p_T(\ell^+\ell^-)>$ 200 GeV and p_T(J)> 100 GeV are considered as merged jet candidates. QCD jets not stemming from boosted Z bosons are excluded from the merged jet analysis by requiring $\tau_{21}<0.5$. The distribution of the pruned mass of the $\ell^+\ell^-$J candidates, depicted in Figure 4.10, presents a good agreement between the observed data and the background expectation.

The sensitivity to a Higgs boson signal is improved at high m_{H} values if the merged jet candidates, $\ell^+\ell^-$J, are preferred over the dijet candidates, $\ell^+\ell^-$jj, when constructing the Higgs boson candidates. At low m_{H}, the particular choice has no effect in the signal sensitivity, but the same criterion is adopted in order to keep the analysis simple.

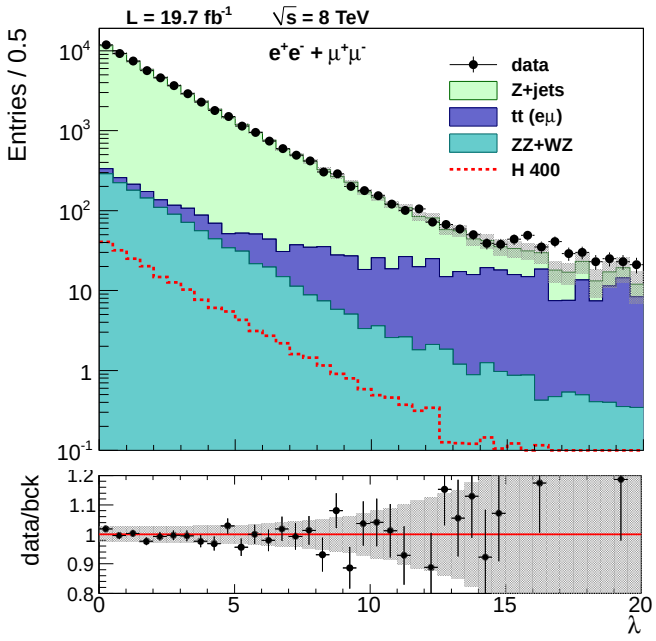


Figure 4.9: Distribution of the $\not\!\!E_T$ significance discriminator, λ, after final selection, except the $\lambda < 10$ condition. Observed data (dots with error bars) are compared to the expected background, together with the systematic uncertainties (filled histograms). The distribution of a Higgs boson signal with m_{H}=400 GeV is also displayed (red dashed histogram). The lower panel depicts the ratio of the observed data and the expected background, together with the systematic uncertainties.

Angular analysis

The decay X→ZZ→$\ell^+\ell^-$q$\bar{\text{q}}$ of an arbitrary particle X is characterized by five decay angles [75] (defined in Figure 4.11 in the H and Z rest frames, assuming X=H): the angle between the proton-proton collision axis (z) and the H → ZZ decay axis (z') in the H rest frame, θ^*; the angle between the zz' plane and the plane of the Z→$\ell^+\ell^-$decay, Φ_1; the angular separation between z' and the negatively charged lepton in the Z rest frame, θ_1; the angle between z' and one of the quarks in the Z rest frame, θ_2; and the angle between the Z→$\ell^+\ell^-$and Z→q$\bar{\text{q}}$ planes, Φ. The shapes of these angular distributions depend on the spin-parity of the particle X. Figure 4.12 depicts normalized distributions of these 5 angles for a SM Higgs boson signal, which is a 0^+ spin-parity state, and for Z+jets events. They are independent of m_{Z} and m_{H}, therefore their different shapes provide an additional handle to isolate signal events in a mass-independent way.

To increase the signal sensitivity, the 5 angles are combined into a linear likeli-

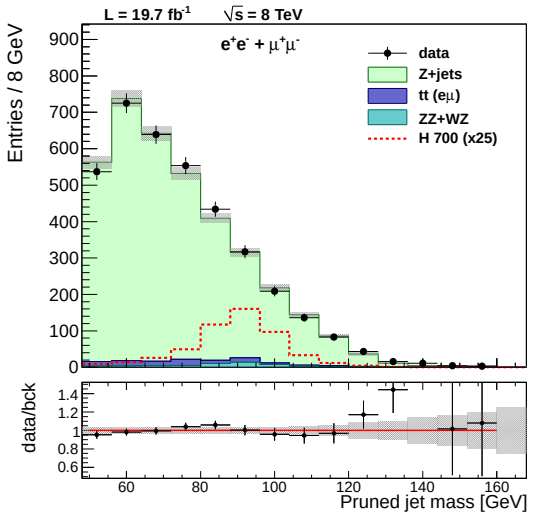


Figure 4.10: Pruned jet mass for the candidate $\ell^+\ell^-$J events, before the event selection. Dots with error bars correspond to observed data, while the background is represented as filled histograms. In addition, the red histogram depicts the distribution for a Higgs boson signal (m_{H}=700 GeV), scaled by a factor 25 for better visibility. The lower panel displays the ratio of the observed data and the expected background, together with the systematic uncertainties.

hood discriminator, LD, using the angular probability distributions for signal and background events, $\mathrm{P_S}(\theta^*, \theta_1, \theta_2, \Phi_1, \Phi)$ and $\mathrm{P_B}(\theta^*, \theta_1, \theta_2, \Phi_1, \Phi)$:

$$\mathrm{LD} = \frac{\mathrm{P_S}}{\mathrm{P_B} + \mathrm{P_S}} \ . \tag{4.1}$$

The $\mathrm{P_S}$ distributions are calculated analytically, $\mathrm{P_S^0}$, and corrected for the detector acceptance (A) as a function of the five angles, assumed to be uncorrelated:

$$\mathrm{P_S} = \mathrm{P_S^0} \cdot A_{\theta^*} \cdot A_{\theta_1} \cdot A_{\theta_2} \cdot A_{\Phi_1} \cdot A_{\Phi}. \tag{4.2}$$

The uncorrelated background probability distributions, $\mathrm{P_B^i}$, are obtained from simulations and, by construction, they already contain acceptance corrections:

$$\mathrm{P_B} = \mathrm{P_B^{\theta^*}} \cdot \mathrm{P_B^{\theta_1}} \cdot \mathrm{P_B^{\theta_2}} \cdot \mathrm{P_B^{\Phi_1}} \cdot \mathrm{P_B^{\Phi}} \ . \tag{4.3}$$

Unlike the background, the signal discriminant distribution peaks at LD values above 0.7 (Figure 4.12, bottom-right).

The angular distributions and the combined angular discriminator are shown in Figure 4.13 for observed data, expected background and a Higgs boson signal be-

fore the event selection. An excellent agreement between the observed data and the background expectations is appreciated, within the systematic uncertainties. Events with low LD values are more likely coming from the Z+jets process. Therefore, the background is cut down by requiring LD>0.5. This requirement is the same for the dijet and merged jet candidates, independent of the Higgs boson mass.

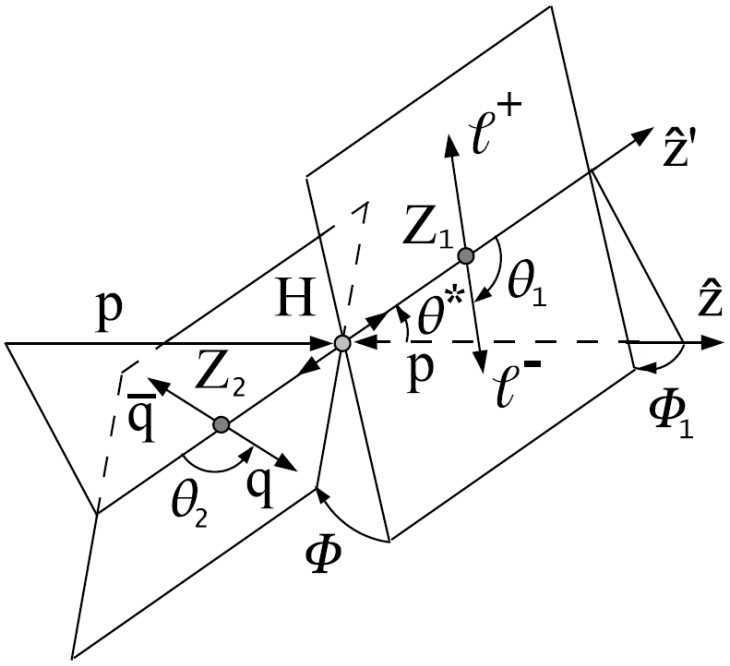


Figure 4.11: Production (θ^*, Φ_1) and decay (θ_1, θ_2, Φ) angles of the H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ process.

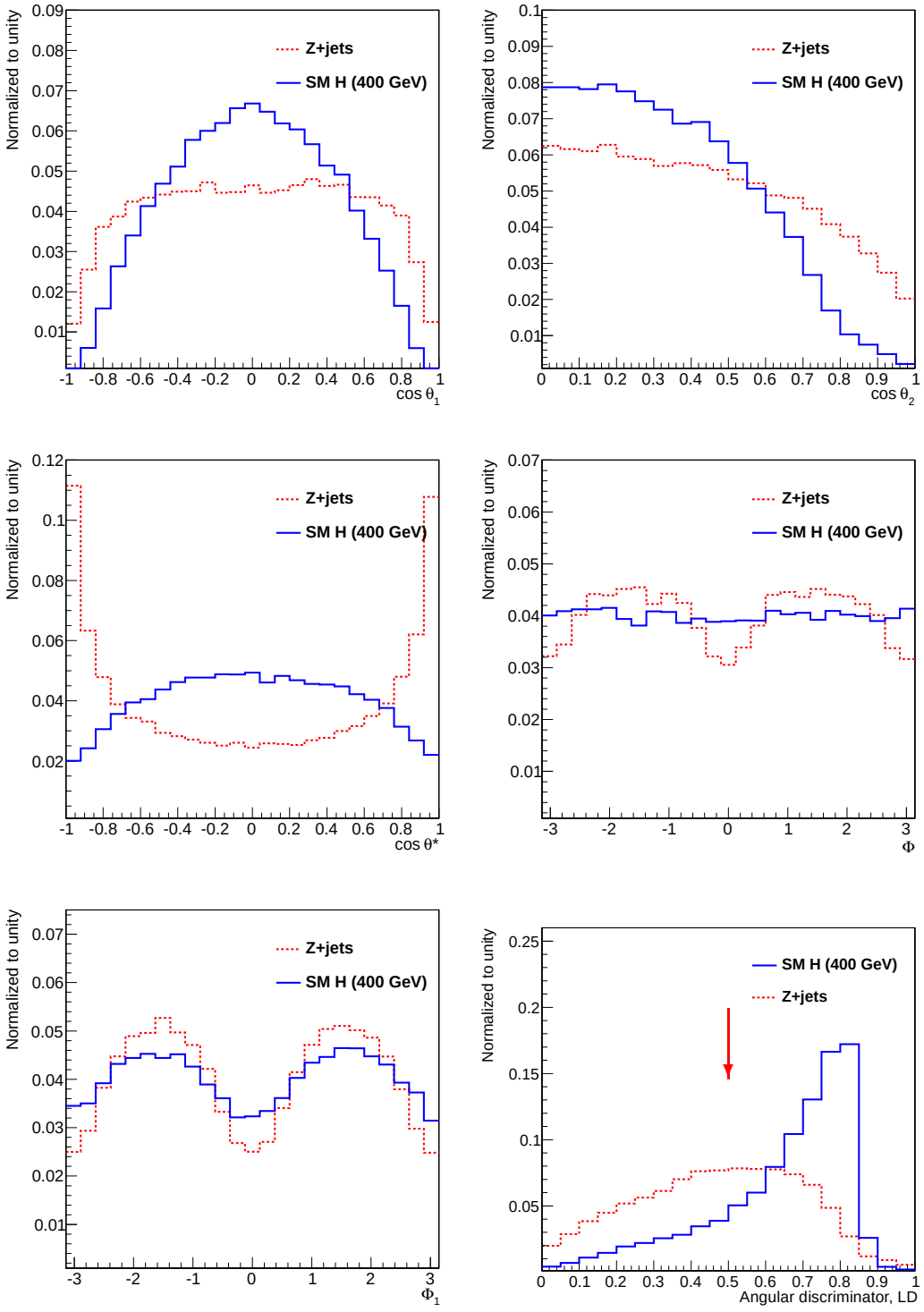


Figure 4.12: Distribution of the production and decay angles for a SM Higgs boson signal with m_{H}=400 GeV (solid blue lines) and for the Z+jets background (dashed red lines) normalized to unity. The bottom-right plot depicts the normalized distribution of the LD discriminator with a red arrow pointing to the selection requirement.

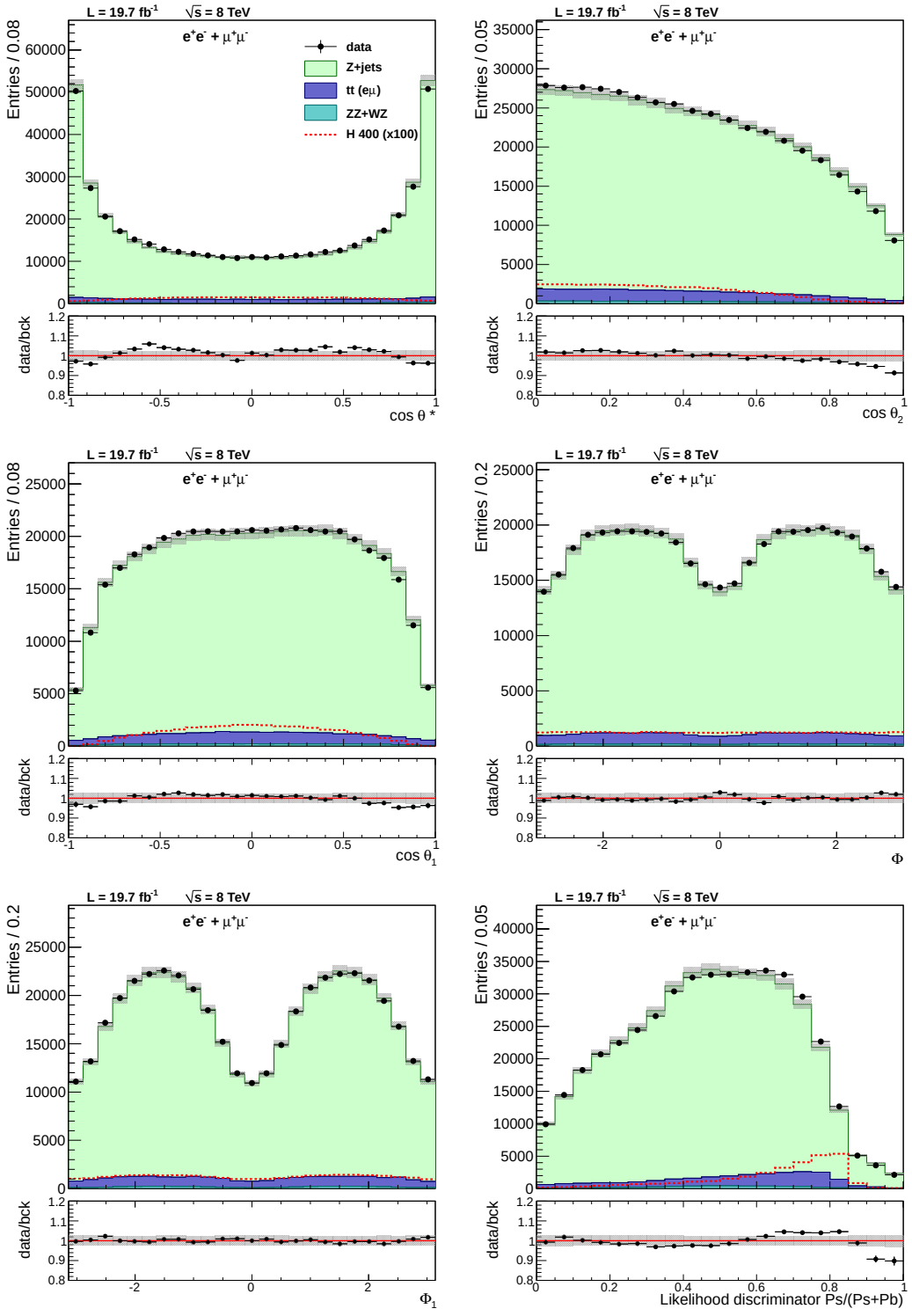


Figure 4.13: Distributions of the decay angles and the likelihood discriminator. Observed data (black dots) are compared to the background expectations (filled histograms). The distribution for a SM Higgs boson with m_{H} =400 GeV is depicted (red histogram), scaled by a factor 100 for better visibility. The lower panel shows the ratio of data to the background expectation. The grey band is the systematic uncertainty.

The selection requirements of the analysis are summarized in Table 4.2. These requirements are chosen to reduce as much as possible the background contamination, enhancing the contribution of the signal, and thus the analysis sensitivity. This can be appreciated in Table 4.3, where the efficiencies of each requirement on observed data, expected background and a Higgs boson signal are listed for the dijet analysis.

Table 4.2: Summary of the selection requirements of the analysis.

Selection cut	Dijet analysis	Merged jet analysis
	Object selection	
$p_T(\ell^{\pm})$	(leading)$>$ 40 GeV and (subleading)$>$ 20 GeV	
$p_T(\ell^{\pm}\ell^{\mp})$	-	$>$ 200 GeV
p_T(AK5 jets)	$>$ 30 GeV	-
p_T(CA8 jets)	-	$>$ 100 GeV
$\|\eta\|(\ell^{\pm})$	$\|\eta\|(e^{\pm}) < 2.5$, $\|\eta\|(\mu^{\pm}) < 2.4$	
$\|\eta\|$(jets)	< 2.4	
β(jets)	> 0.2	
ΔR	$\Delta R(\ell$,AK5jet)>0.5	$\Delta R(\ell$,CA8jet)>0.8
	Event selection	
m_{jj}	$\in$ [71, 111] GeV	-
m_J	-	$\in$ [71, 111] GeV
$m_{\ell\ell}$	$\in$ [76, 106] GeV	
LD	>0.5	
λ	<10	
τ_{21}	-	< 0.5

Production mechanism

The contribution of the vector boson fusion production mechanism to the SM Higgs boson cross section is about 10% at the lower masses considered in the analysis, m_{H}=230 GeV, and increases to around 50% at m_{H}=1 TeV, providing a similar contribution than the gluon fusion production mechanism (Figure 1.2 in section 1.2).

The VBF mechanism is interesting for several reasons. First of all, the SM accurately predicts the production cross section for the ggH and VBF diagrams. If a Higgs boson is found at a certain mass, measuring the different production cross sections is essential to test the predictions. Also, studying the VBF production allows to search

Table 4.3: Efficiencies of the selection requirements for observed data, expected backgrounds and a hypothetical Higgs boson with m_{H}=300 GeV. The relative efficiencies of each selection with respect to the previous condition are denoted as Rel., while the absolute efficiency is labeled as Abs.

	Data		H (300 GeV)		Z+jets		Diboson		$t\bar{t}$	
	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.
Obj. sel.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$m_{\ell\ell}$	0.77	0.77	0.88	0.88	0.86	0.86	0.86	0.86	0.19	0.19
LD	0.46	0.35	0.75	0.66	0.48	0.41	0.53	0.45	0.61	0.12
λ	0.96	0.34	0.99	0.65	0.99	0.41	0.96	0.43	0.43	0.05
m_{jj}	0.39	0.13	0.60	0.39	0.36	0.15	0.61	0.27	0.35	0.02

for Higgs bosons produced in models with enhanced production modes. For example, fermiophobic models contain Higgs bosons that do not couple to fermions, and thus, the gluon fusion mechanism is suppressed. In that kind of models the VBF mechanism is the dominant production mode for heavy Higgs bosons in a hadron collider.

The analysis divides the selected signal sample in two categories, one of them dominated by the gluon fusion production and the other enhancing VBF signals. This separation exploits the different topologies presented by both production mechanisms. The gluon fusion mechanism consists of the interaction of two gluons of the colliding protons, through a heavy quark loop. In the VBF mechanism, the Higgs bosons are produced by quark scattering, with weak gauge bosons as intermediate particles. The quarks emitting the weak gauge bosons hadronize, yielding jets in the direction of the incoming protons. The events are classified as VBF or ggH by identifying the presence of such additional jets.

To study the aforementioned forward jets, all the reconstructed AK5 jets within $|\eta| = 4.7$ are considered, excluding all the jets already identified as hadronic Z boson decays, either resolved or merged. Outside the tracker acceptance, $|\eta|$>2.4, jets are reconstructed using information of the forward hadron calorimeter. In this context, the pile-up mitigation variable β is no longer valid (Equation 3.3), since it uses tracking information to discriminate particles coming from the primary vertex and constituents from pile-up vertices. For these additional jets, a pile-up multivariate-based discriminator is used, taking as input jet shape variables [76].

In gluon fusion events, forward additional jets may be present, coming from pile-up or from the remnants of the hard interaction (underlying event). Figure 4.14 shows the η distribution of the two highest p_T additional jets (AK5) for a Higgs boson

signal in the ggH and VBF production mechanisms. The ggH additional jets are mainly located in the central region of the detector, while the VBF additional jets are primarily in the forward direction. Moreover, the additional jet pair produced by the VBF mechanism has a higher mass than the random pairs of additional jets in ggH events. In summary, VBF processes are identified as those events containing an additional high mass jet pair, m_{jj}> 500 GeV, with large η separation, $\Delta\eta > 3.5$ (Figure 4.15).

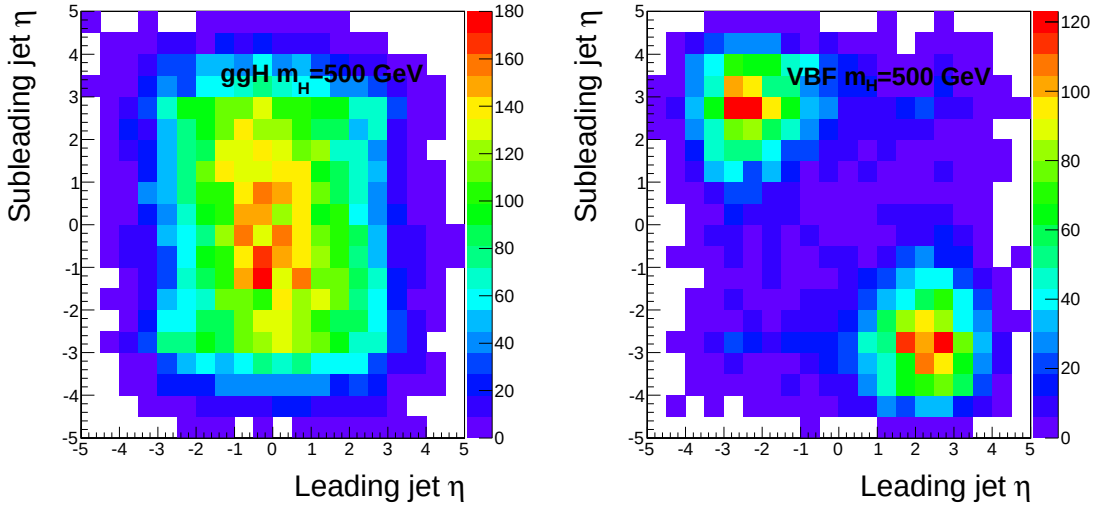


Figure 4.14: Distribution of η of the p_T-leading and p_T-subleading jets for a SM Higgs boson with m_{H}=500 GeV. The two main production mechanisms are depicted: gluon fusion (left) and vector boson fusion (right).

Simulated samples are used to estimate the purity of the VBF event selection. The rate of ggH events in the VBF sample is around 10%. Conversely, about 30% of the events in the ggH sample are VBF events. These figures are taken into account in the statistical procedure that quantifies the presence of hypothetical signals, predicted by different models, in the observed data.

Event categories

In order to significantly increase the sensitivity to hypothetical Higgs boson signals, the selected events are classified into exclusive sub-samples with different signal-to-background ratios. Fourteen different categories are defined: two categories of VBF events, split by lepton flavour (e^+e^-, $\mu^+\mu^-$); and twelve categories of ggH events: 6

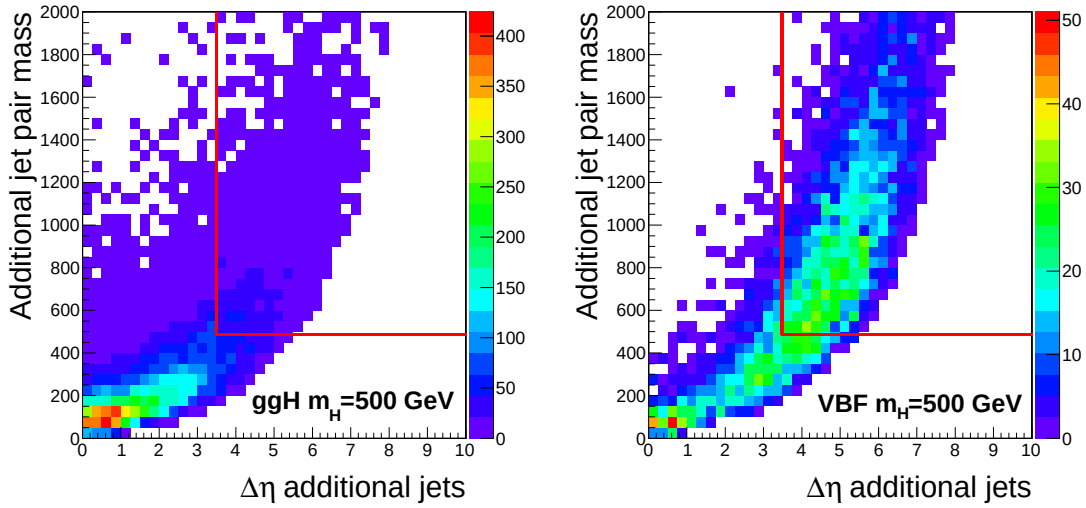


Figure 4.15: Invariant mass versus η difference of the p_T-leading and p_T-subleading jets for a SM Higgs boson with $m_{\text{H}} = 500$ GeV. The distribution is used to distinguish events produced by the vector boson fusion (right) and the gluon fusion (left). The red lines delimit the VBF selection (m_{jj}> 500 GeV and $\Delta\eta > 3.5$)

dijet and 6 merged jet categories, each split by the lepton flavour and by the number of b-tagged (sub-)jets (0 b-tag, 1 b-tag and 2 b-tag). The number of observed events after the selection in each of these sub-samples is listed in Table 4.4

Table 4.4: Number of observed events in each of the 14 exclusive categories.

	VBF	Merged jet			Dijet		
		0 b-tag	1 b-tag	2 b-tag	0 b-tag	1 b-tag	2 b-tag
$\mu^+\mu^-$q$\bar{\text{q}}$	276	201	76	7	15806	5253	467
e^+e^-q$\bar{\text{q}}$	213	175	73	3	13173	4227	381

Only one Higgs boson candidate is selected per event. If more than one candidate passes the selection requirements, the following criteria are used to select the best candidate: first, VBF events are preferred over ggH events. In case of further ambiguity, the candidates containing a boosted hadronic Z boson ($\ell^+\ell^-$J) are preferred over the dijet candidates ($\ell^+\ell^-$q$\bar{\text{q}}$). In case that multiple candidates belonging to the same category pass all the selection requirements, the candidate with the highest number of b-tagged (sub-)jets is selected. Finally, if the number of b-tags is the same, the candidate minimizing the quantity $|m_{jj} - m_{\text{Z}}| + |m_{\ell^+\ell^-} - m_{\text{Z}}|$ is considered as the Higgs boson candidate.

4.6 Background determination

The three main background sources in this analysis are, in order of importance, events from Z+jets, $t\bar{t}$ and diboson production. Other contributions such as W+jets or multi-jet events, containing jets misidentified as leptons, are found to be negligible after the tight dilepton selection. Since the Higgs boson signal is several orders of magnitude smaller than the contamination, the proper understanding of the backgrounds is the essential part of the analysis. This section describes the methods used to estimate and validate the different background contributions.

Z+jets background

The main background is composed by events with Z bosons decaying leptonically, produced in association with one or more hadronic jets. It is estimated using large simulated samples, which include several special samples (section 4.3).

The simulation of the Z+jets background is validated using the m_{jj} control region. All the relevant observables of the analysis are studied in this region, revealing an overall good agreement between the observed data and the simulation. Nevertheless, some small discrepancies are found in the transverse momentum distributions of the $\ell^+\ell^-$and $\ell^+\ell^-\mathrm{q\bar{q}}$ systems. This mismodelling appears simultaneously in the signal and control regions (Figure 4.16, left), hence the simulation is corrected using event-by-event weights determined in the control region.

The weights are calculated to match the $p_T(\ell^+\ell^-\mathrm{jj})$ distribution of the simulated Z+jets events to that of the observed data. They are extracted from a fit to the ratio of the $p_T(\ell^+\ell^-\mathrm{jj})$ ditributions in the observed data and in the simulation for the control region. The function used to perform the fit,

$$f(p_T^{\ell\ell\mathrm{jj}}) = \left(1 + \frac{1}{a + bp_T^2}\right) \frac{1}{e^{-p_T^2/c} + 1} \ , \tag{4.4}$$

depends on three parameters (a, b and c), which are determined by the fitting procedure. The function grows for low momentum values and tends to 1 at large $p_T(\ell^+\ell^-\mathrm{jj})$ values, where the simulation correctly reproduces the observed momentum spectra (Figure 4.16, right). The correction of the p_T distribution in the simulation is about -20% at $p_T(\ell^+\ell^-\mathrm{jj})$=0, grows to 9% around $p_T(\ell^+\ell^-\mathrm{jj})$=70 GeV, and slowly decreases towards higher p_T values.

Similarly, the transverse momentum of the $\ell^+\ell^-$ system is also corrected at very high momentum values, affecting particularly the merged jet analysis. A linear fit

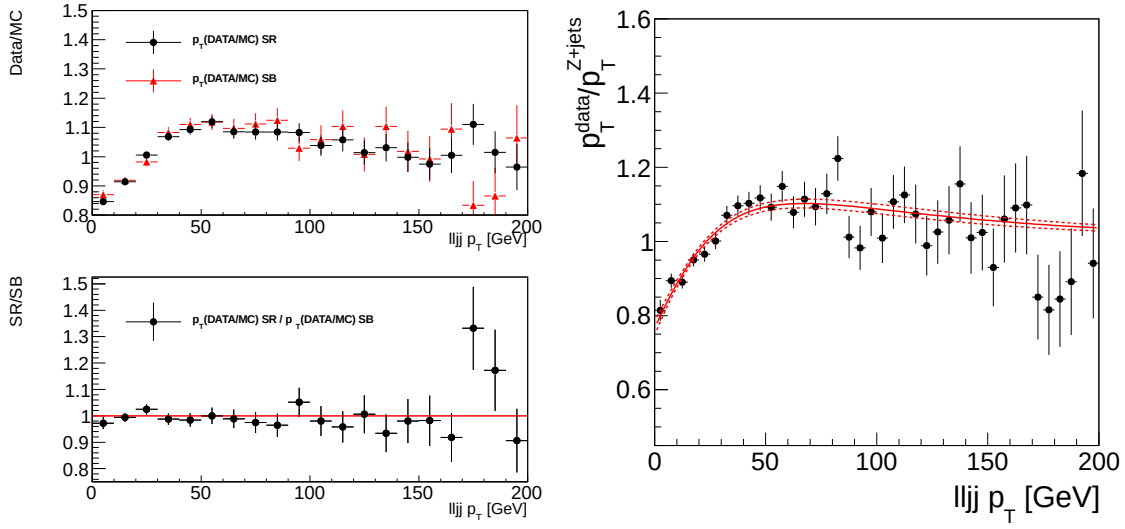


Figure 4.16: Left: the upper panel shows the ratio of $p_T(\ell^+\ell^-\text{jj})$ distributions in observed data and in the simulation in the control (red triangles) and the signal regions (black dots). The lower panel plot shows the division of the ratios above. Right: ratio of the $p_T(\ell^+\ell^-\text{jj})$ distribution in observed data and in the background in the control region. The solid red line represents the fit used to extract the weights and the red dashed lines denote the one standard deviations ($\pm 1\sigma$) of the fit.

to the ratio of the observed data over the simulation in the control region is used to extract an event-by-event weight that is applied to the Z+jets Madgraph simulation (Figure 4.17). The correction is small for low $p_T(\ell^+\ell^-)$ values and is 70% at high dilepton tranverse momenta. The effect of these two corrections on the invariant mass of the $\ell^+\ell^-$q$\bar{\text{q}}$ system is small, below 3%.

The normalization of the Z+jets sample in the signal region is constrained to the relative normalization of the observed data and the simulation in the control region, independently on each of the 14 categories.

Background from t$\bar{\text{t}}$ events

The process t$\bar{\text{t}}$ → W$^+$W$^-$b$\bar{\text{b}}$ → $\ell^+\nu_\ell\ell'^-\bar{\nu}_\ell$b$\bar{\text{b}}$ is an important source of contamination in the categories containing two b jets. The top quark pair production is estimated taking advantage of the fact that the leptons coming from the W boson decays can be of different flavour. The estimation uses a sample of observed data that satisfies the requirements of the analysis (detailed in section 4.5), except that events with

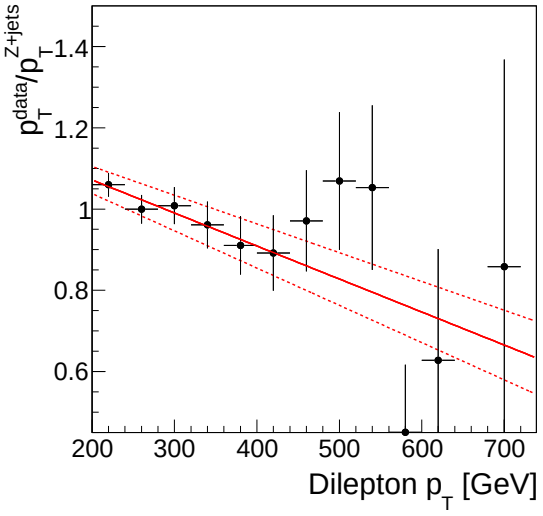


Figure 4.17: Ratio of the $p_T(\ell^+\ell^-)$ distributions in the observed data over the simulation for the merged-jet events. The solid red line represents the fit used to extract the weights and the dashed red lines denote the one standard deviations ($\pm 1\sigma$) of the fit.

opposite-flavour leptons are selected ($e^{\pm}\mu^{\mp}$), instead of e^+e^- or $\mu^+\mu^-$. Unlike the $\ell^+\ell^-$jj sample, dominated by Z+jets events, the $e^{\pm}\mu^{\mp}$jj sample consists mostly of $t\bar{t}$ events, as clearly visible in the reconstructed invariant mass distribution of the $\ell^+\ell^-$jj system for the selected $e^{\pm}\mu^{\mp}$ events (Figure 4.18, left).

The sample of $t\bar{t}$ events decaying to $e^+\mu^-$ or μ^+e^- behaves as the sample with $t\bar{t}$ events decaying to e^+e^- or $\mu^+\mu^-$. Therefore, the $e^{\pm}\mu^{\mp}$data sample is used as an estimate of the $t\bar{t}$ background contribution. This assumption is tested with the observed data using a top-enriched sample. In order to select a pure sample of $t\bar{t}$, the events are required to be outside the dilepton mass window considered in the analysis, $m_{\ell\ell} \notin [76,106]$ GeV. The remaining non-top background contribution is eliminated selecting events with significant $\not\!\!E_T$, by requiring $\lambda > 8$. The comparison between the $e^+\mu^- + \mu^+e^-$ and $e^+\mu^- + \mu^+e^-$ samples with these requirements is displayed in the right plot of Figure 4.18. The very good agreement, both in shape and normalization, of the $m_{\ell^+\ell^-q\bar{q}}$ distributions confirms that the opposite-lepton-flavour data sample gives a reliable estimation of the $t\bar{t}$ background in the analysis, justifying the use of the $e^{\pm}\mu^{\mp}$sample as an estimator of the $t\bar{t}$ background in the signal region.

Other backgrounds like Z+jets$\rightarrow\tau^+\tau^-$+jets, W^+W^- or the single-top production, in which the opposite-lepton-flavour and same-flavor components are equivalent, are

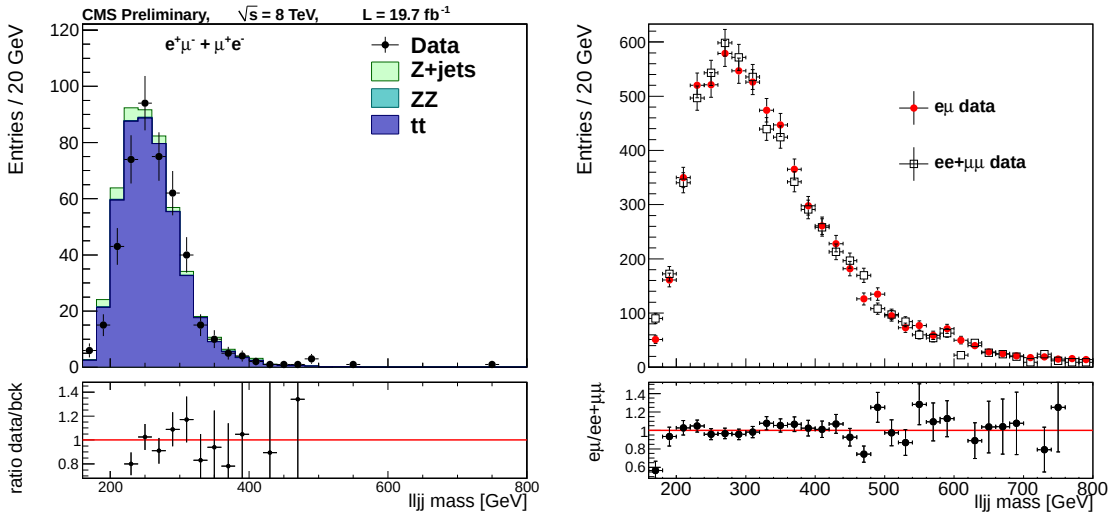


Figure 4.18: Left: invariant mass of the $\ell^+\ell^-$jj system for selected $e^{\pm}\mu^{\mp}$events. Black dots with error bars correspond to the observed data, while the background contributions (dominated by $t\bar{t}$) are represented as filled histograms. Right: Comparison of the reconstructed invariant mass for $e^+e^-+\mu^+\mu^-$ (open boxes) and $e^+\mu^-+e^-\mu^+$ (red dots) events in a top-enriched sample ($\lambda > 8$, $m_{\ell\ell}\notin$[76,106] GeV). The lower panels present the ratio of observed data to expected background (left) and the ratio of eμ data to ee+$\mu\mu$ data (right).

also included in the data-driven $e^{\pm}\mu^{\mp}$ estimation. These backgrounds are found to be a minor contribution.

In the VBF categories, the $t\bar{t}$ background is almost negligible. Due to the low yields of the data-driven method, the statistical significance of the sample is not enough to give an accurate prediction. Therefore, in this case the top quark pair production is estimated directly from the simulation.

Diboson background

The diboson contamination (ZZ, ZW and WW) is almost negligible in comparison to the dominant Z+jets background, due to its relatively small production cross section, but still relevant for the Higgs boson search. It is accurately estimated using a large number of simulated events.

4.7 Systematic uncertainties

The systematic uncertainties affect both the shape and normalization of the Higgs boson signal and background estimations. It is extremely important to correctly account for the uncertainties when comparing the observed data to the predictions. Several systematic effects have been studied like theoretical uncertainties, experimental-related systematics (detector efficiencies, resolutions, calibrations, etc.), and the uncertainties derived from the background estimation methods.

This section describes these uncertainties, the method used to extract them and their estimated values. Tables 4.9 and 4.10 summarize the signal and background systematic uncertainties.

Common systematic uncertainties

The experimental-related effects affect similarly the background and signal estimations, such as the lepton reconstruction efficiencies, the jet energy determination or the b-tagging corrections. The most relevant systematics influencing both signal and background are described below.

The dominant systematic effect influencing the leptons is the uncertainty in the detection efficiency. It is estimated as the uncertainty in the leptonic scale factors (SF) applied to correct the simulated samples, as described in section 3.2. For the muon detection efficiency, three effects are added up in quadrature: identification (0.5%), isolation (0.2%) and trigger (1.7%) uncertaintes. The result amounts to 1.8%. For the electron detection efficiency, the systematic uncertainty is largely dominated by the 2% identification systematic uncertainty.

The jet energy scale systematic, which comes from the non-linear response of the HCAL in the energy measurement, is the dominant uncertainty. For each jet, a correction is applied to map the measured jet energy to its true energy. This correction is p_T and η dependent, and is determined using a γ^*/Z+jets data sample [52]. The uncertainty derived from this correction is used to evaluate the jet energy scale systematic, repeating the analysis with jet scales varied up and down by one standard deviation. For the Z+jets background, two alternate shapes are derived (Figure 4.19). It amounts to a 3% systematic uncertainty for the dijet analysis, a 12% systematic in the VBF analysis and a 0.9% systematic in the merged jet analysis. For the signal, the shape variation is small, and thus the difference of efficiency is taken as a normalization systematic uncertainty, as listed for a few m_{H} hypotheses in Table 4.5.

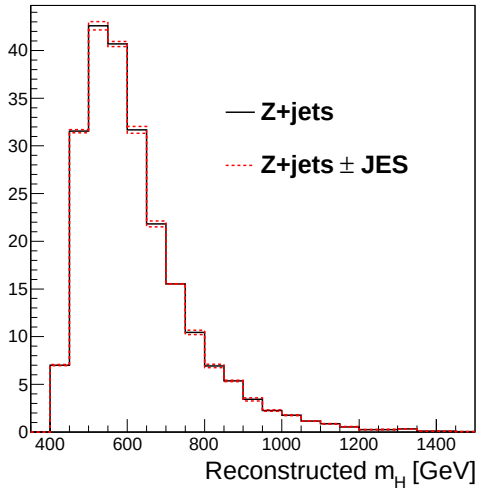


Figure 4.19: Evaluation of the jet energy scale systematic uncertainty in the merged jet analysis. The reconstructed m_{H} of the $\ell^+\ell^-$J candidates in the Z+jets background (solid black line) is shown together with two alternate shapes produced by changing the jet energy scale by one standard deviation (dashed red lines).

Table 4.5: Jet energy scale systematic uncertainty (in %) for several Higgs boson mass hypotheses.

Mass [GeV]	Dijet	Merged jet	VBF
250	4.2	-	9.2
400	1.2	0.2	8.6
600	0.9	0.2	8.7
800	0.7	0.1	8.6

The uncertainty in the quark flavour identification is one of the main systematic uncertainties of the ggH categories. Two effects are studied, the uncertainties in the b-tagging efficiency and the uncertainties in the misidentification of light jets as b jets. They are estimated by varying the corresponding scale factors, SF_{b} and SF_{light}, by one standard deviation, in the two operational working points used in the analysis (JPL, JPM). The change of efficiency when varying the scale factors gives the systematic uncertainty. Table 4.6 shows the uncertainties for the b-tagging efficiency and mistag rates computed for the Z+jets background and for several signals.

The dijet and merged jet categories, separated into the number of b-tagged jets, are listed in the table.

Table 4.6: b-tagging and mistagging scale factor systematic uncertainties (in %) for the background and for several Higgs boson signal hypotheses.

	Dijet analysis			Merged jet analysis		
	0-btag	1-btag	2-btag	0-btag	1-btag	2-btag
b-tagging efficiency systematic						
Z+jets	0.4	0.8	4.6	0.3	0.7	4.5
m_{H}= 250 GeV	2.1	3.8	5.5	-	-	-
m_{H}= 400 GeV	3.9	4.5	5.7	0.8	0.1	5.4
m_{H}= 600 GeV	4.4	5.2	6.1	0.7	0.1	5.6
m_{H}= 1000 GeV	-	-	-	1.1	0.2	7.5
Mistagging rate systematic						
Z+jets	1.1	3.7	3.6	1.3	5.0	6.5
m_{H}= 250 GeV	1.0	2.1	0.6	-	-	-
m_{H}= 400 GeV	1.2	2.4	0.5	1.3	2.8	0.4
m_{H}= 600 GeV	1.2	2.7	0.6	1.4	3.0	0.3
m_{H}= 1000 GeV	-	-	-	2.1	4.8	1.3

Another source of uncertainty comes from the matching in the simulation to the number of interactions per bunch crossing in the data, as explained in section 4.2. This uncertainty is evaluated by varying the inelastic proton-proton cross section by ±5% (Figure 4.4). The effect on the background yields is taken as a normalization systematic uncertainty. The Table 4.7 summarizes these systematic uncertainties for the main background and for the signal.

The instantaneous luminosity is measured in CMS using the silicon pixel detector, the closest detector to the beam pipe. The method relies in the fact that the instantaneous luminosity is proportional to the number of reconstructed clusters in the pixel detector in each bunch crossing. The total uncertainty on the luminosity determination is calculated to be 2.6% [64], and is treated as a systematic uncertainty on the signal normalization.

The systematic uncertainty due to the $\not\!\!E_T$ treatment is covered by other uncertainties, such as the jet reconstruction and the pile-up systematic uncertainties. An additional systematic uncertainty is considered due to the $\not\!\!E_T$ significance scaling, described in section 4.5. The difference in efficiency when not applying the λ scaling

Table 4.7: Pile-up systematic uncertainties (in %) for the Z+jets background and for several Higgs boson signals.

	Dijet	Merged jet	VBF
Z+jets	1.0	0.7	0.1
m_{H}=250 GeV	0.3	-	1.5
m_{H}=400 GeV	0.2	0.2	1.3
m_{H}=600 GeV	0.1	0.07	0.5
m_{H}=800 GeV	0.1	0.3	1.1

is used as a normalization systematic uncertainty.

The performance of the boosted gauge boson identification using the τ_{21} variable has been analyzed on dedicated studies [77]. This tagging dicriminator is tested using a t$\bar{\text{t}}$ control sample, containing a boosted W boson. By identifying that boson at generator level in the simulation and comparing it with the efficiency when applying the $\tau_{21} < 0.5$ requirement, a systematic uncertainty of 10% is assigned to the signal and diboson events containing merged jets.

Signal systematic uncertainties

The cross section for the gluon fusion and vector boson fusion mechanisms are computed using QCD calculations at NNLO precision and electroweak calculations at NLO precision [23]. A normalization systematic uncertainty in these cross sections covers the uncertainty due to higher order terms missing in the QCD and electroweak corrections on the perturbation series [22]. For the gluon fusion mechanism it amounts to 10% at m_{H}=250 GeV and increases with the mass up to 14% at m_{H}=1 TeV. For the VBF mechanism, the uncertainty is smaller, around 3% at low masses and up to 6% at high mass hypotheses. These uncertainties are only relevant to the measurement when compared to the SM expectation and do not affect the absolute cross section measurement.

Another source of uncertainty comes from the underlying parton distribution functions (PDF) used to produce the simulated signals. A change in the underlying PDFs modifies the kinematics of the Higgs boson production. This uncertainty is propagated into the analysis by evaluating the signal efficiency using three PDF sets: CTEQ6.6 [78], MSTW08 [79] and NNPDF2.0 [80]. Each of these PDF parametrizations uses distinct values for the strong coupling, α_s, and treats the heavy quarks

differently. The total normalization systematic uncertainty is taken as the envelope of the three results [81]. The PDF systematic uncertainty is less than 1% at m_{H}=200 GeV and increases with the mass, up to 4% at 1 TeV. Table 4.8 summarizes the PDF uncertainty for several signal hypotheses.

Table 4.8: PDF+α_s systematic uncertainties (in %) for several Higgs boson signals.

Mass [GeV]	CTEQ6.6	MSTW08	NNPDF2.0	Total
200	0.7	0.5	0.8	0.8
400	1.0	0.6	1.4	1.4
600	1.1	0.8	1.5	1.5
800	2.0	1.5	2.7	2.7

The modelling of the Higgs boson signals, described in section 4.4, is another source of systematic uncertainty. This modelling incorporates a Higgs boson lineshape using the complex-pole scheme correction and including the gg→ZZ interference. Two alternate shapes to the invariant mass distribution are considered as the systematic uncertainty associated to these corrections. The original lineshape (without CPS and interference) and its symmetrization are taken as the shape systematic uncertainties due to the signal modelling. Figure 4.20 displays those alternative shapes in the dijet analysis for a Higgs boson hypothesis of m_{H}=550 GeV. This uncertainty is negligible for Higgs boson masses below 400 GeV, where the complex-pole scheme is a subdominant effect, affecting predominantly the higher masses. For instance, it changes the signal acceptance around 7% for very high Higgs masses.

Background systematic uncertainties

The uncertainty due to the Z+jets modelling is extracted from the weights used to correct the simulations, described in section 4.6. The one standard deviations ($\pm1\sigma$) in the fits performed to extract the weights are used to estimate two alternate shapes in the distribution of the $\ell^+\ell^-\mathrm{q}\bar{\mathrm{q}}$ reconstructed invariant mass. An example of the alternate Z+jets shapes is displayed in Figure 4.21.

The diboson backgrounds ZZ and WZ are estimated using the simulations. A systematic uncertainty on the normalization is estimated by the uncertainty in the production cross section measured by CMS, which amounts to 12% [82].

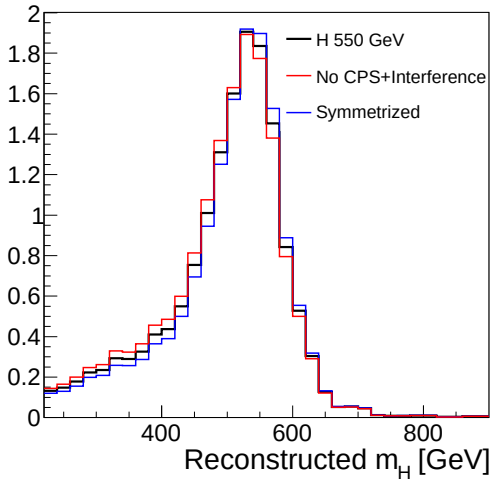


Figure 4.20: Estimation of the systematic uncertainty due to the Higgs boson lineshape modelling in the $\ell^+\ell^-$jj analysis. The reconstructed m_{H} mass for a Higgs boson hypothesis with m_{H}=550 GeV (black) and two alternate shapes constructed by not considering the CPS approximation and gg→ZZ interference (red) and its symmetrization (blue).

Table 4.10 summarizes the systematic uncertainties affecting the background estimations.

Table 4.9: Summary of systematic uncertainties (in %) on the signal expectation.

Source	VBF analysis	Merged jet analysis	Dijet analysis
Muon trigger, isolation & ID	1.8		
Electron trigger isolation & ID	2.0		
Electron energy scale	-	0.1	0.2
Muon momentum scale	-	0.1	0.1
Jet reconstruction	1-13	0.2-0.6	1-4
b-tagging efficiency and mistag rate	-	1.4-7.5	2.1-6.1
$\lambda(\not{E}_T)$	-	-	0.6
Boosted Z tagging	10 (merged jets only)		-
Pile-up	0.1-2		
PDFs	0.8-8		
Lineshape	1-7		
Luminosity	2.6		
ggH cross-section	9-14		
VBF cross-section	3-6		

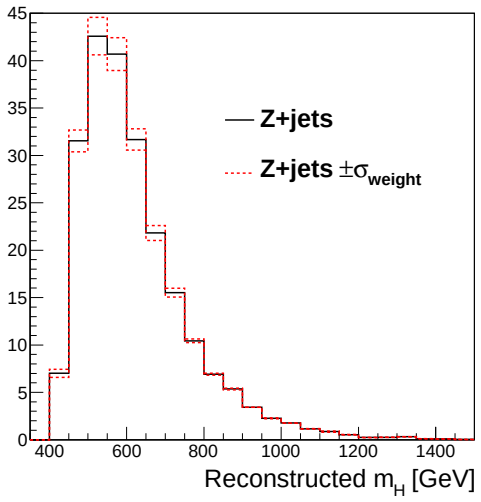


Figure 4.21: Estimation of the Z+jets systematic uncertainty due to the p_T(ZZ)-dependent weights. The Z+jets reconstructed $\ell^+\ell^-$J mass (solid black line) and two alternate shapes produced varying the $p_T(\ell^+\ell^-\text{J})$ fits by one standard deviation (dashed red lines) are depicted.

Table 4.10: Summary of systematic uncertainties (in %) on the background determination.

Source		VBF analysis	Merged jet analysis	Dijet analysis
Muon trigger & ID		1.8		
Muon momentum scale		-	0.2	0.1
Electron trigger & ID		2.0		
Electron energy scale		-	0.8	0.1
Jet energy scale		12-19	0.9	3.6
b-tagging efficiency	0-btag	-	0.3	0.4
	1-btag	-	0.7	0.8
	2-btag	-	4.5	4.6
Mistag rate	0-btag	-	1.4	1.1
	1-btag	-	5.0	3.7
	2-btag	-	6.5	3.6
$\not\!\!E_T$ significance, λ		-	-	0.3
Pile-up		1	0.7	0.1
Z+jets p_T(ZZ) correction		-	0.8	3
Diboson cross section		12		
Luminosity		2.6		

Chapter 5

Results and interpretation

The $m_{\ell\ell qq}$ distribution of the background is exponentially falling, in contrast with the resonant shape of the hypothetical Higgs boson signals. Advanced statistical methods exploit this singular behaviour of the mass distribution to quantify the compatibility of the observed data with the background expectations. Localized deviations from the predictions are interpreted as a signal coming from either the SM or BSM theories. This chapter presents the statistical interpretation of the results of the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ analysis, in the context of the search for a heavy SM Higgs boson and for an electroweak singlet particle.

5.1 Statistical data analysis

The number of events selected in the analysis described in the previous chapter are listed in Table 5.1 and depicted in Figure 5.1. The three main categories, low-mass dijet, high-mass merged jet and VBF, split into b-tag categories are shown, with the lepton categories summed up together. The events observed in data are compared to the number of expected background events from $e^{\pm}\mu^{\mp}$ data, and Z+jets and diboson simulations. Finally, the yields for three different Higgs boson mass hypotheses, 300 GeV (low mass), 600 GeV (intermediate mass) and 900 GeV (high mass), are also displayed. No statistically significant deviation is found in the overall number of observed data when compared with the background expectations in any of the 14 exclusive categories.

The distributions of the statistical discriminator, $m_{\ell\ell q\bar{q}}$, are displayed for the observed data, the expected backgrounds and a hypothetical signal for the 6 low-mass dijet categories (Figure 5.2), for the 6 high-mass merged jet categories (Figure 5.3) and for the 2 VBF categories (Figure 5.4), separately for the di-electron and di-muon channels. The observed data show a good agreement with the expected background within the uncertanties in the 14 exclusive categories, with no evidence of clearly

peaking structures over the whole mass range.

Table 5.1: Expected numbers of background events and observed yields in data. Signal expectations for m_{H}=300, 600 and 900 GeV are also listed.

	VBF analysis	Merged jet analysis			Dijet analysis		
		0 b-tag	1 b-tag	2 b-tag	0 b-tag	1 b-tag	2 b-tag
Background	485	381	133	16	29027	9451	823
Data	489	376	149	10	28979	9480	848
m_{H}[GeV]	Signal expectation						
300	19.7	negligible			229.5	110.6	32.7
600	6.0	24.3	13.9	6.1	15.8	8.1	3.2
900	1.9	4.3	3.0	1.4	negligible		

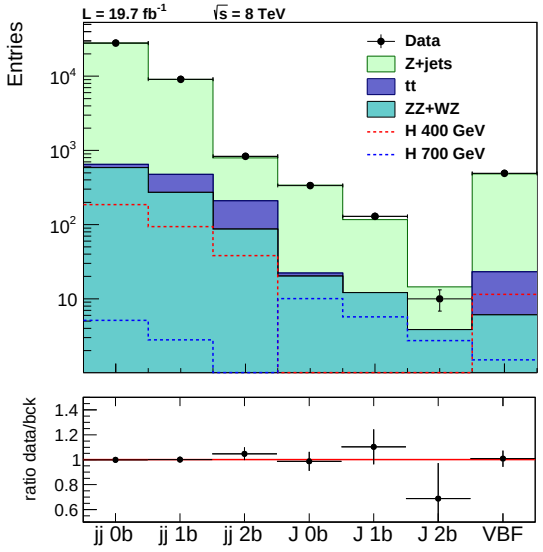


Figure 5.1: Numbers of events in each of the exclusive categories for the observed data (dots) and the estimated background (filled histograms) for $\mu^+\mu^-$ and $\mathrm{e}^+\mathrm{e}^-$ combined. Two signal hypotheses are included as dashed histograms, for 400 GeV and 700 GeV mass. The lower panel shows the ratio of the observed data and the expected background.

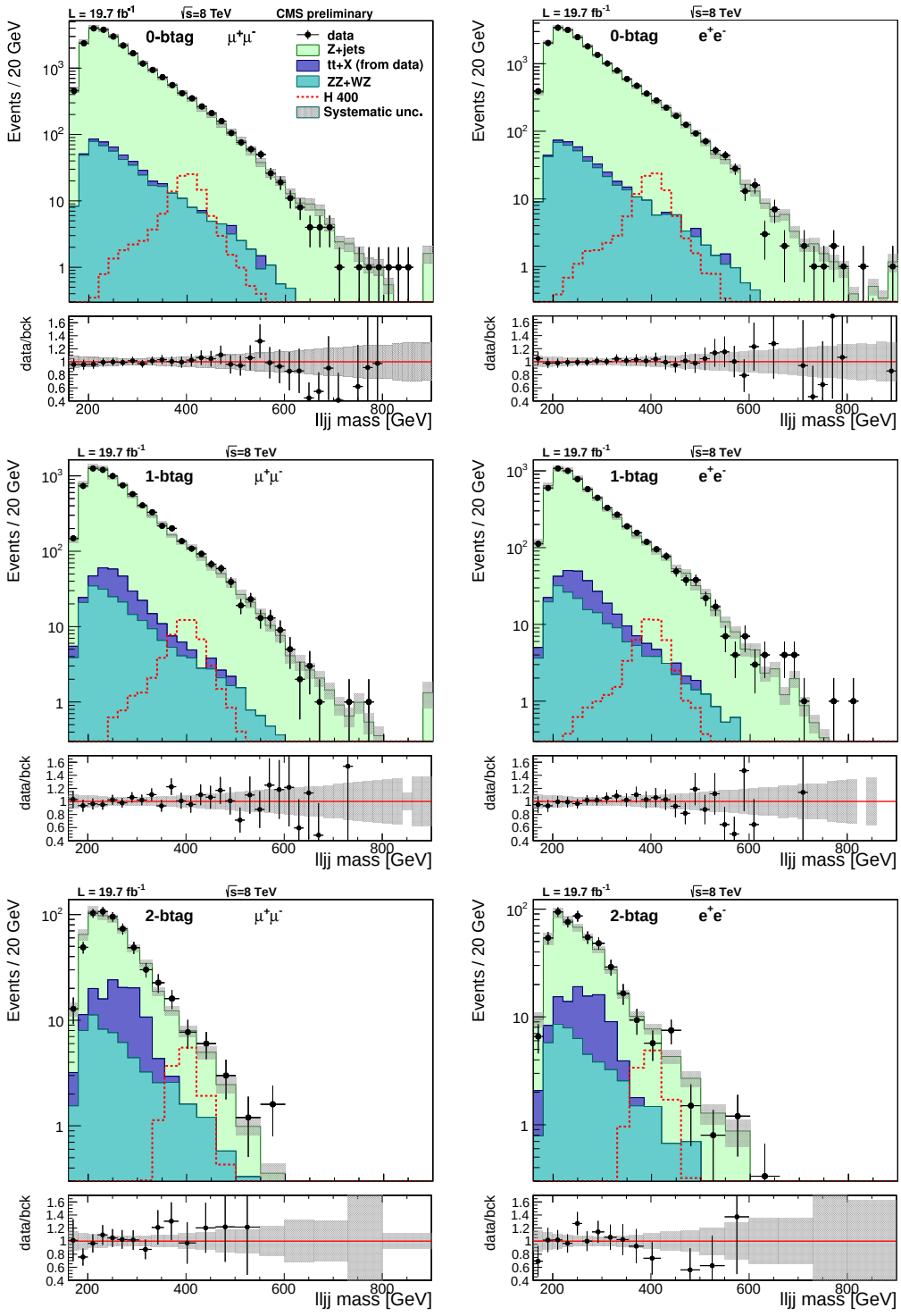


Figure 5.2: Distribution of $m_{\ell\ell q\bar{q}}$ distributions for the dijet analysis after final selection in the six categories: muon (left column) and electron (right column) channels; 0-btag (top), 1-btag (middle) and 2-btag (bottom). Dots with error bars denote the observed data, the histograms represent the expected backgrounds, and the grey band correspond to the systematic uncertainties. A 400 GeV SM Higgs boson is also shown (red dashed line). The lower panels show the ratio of the observed data and the expected background.

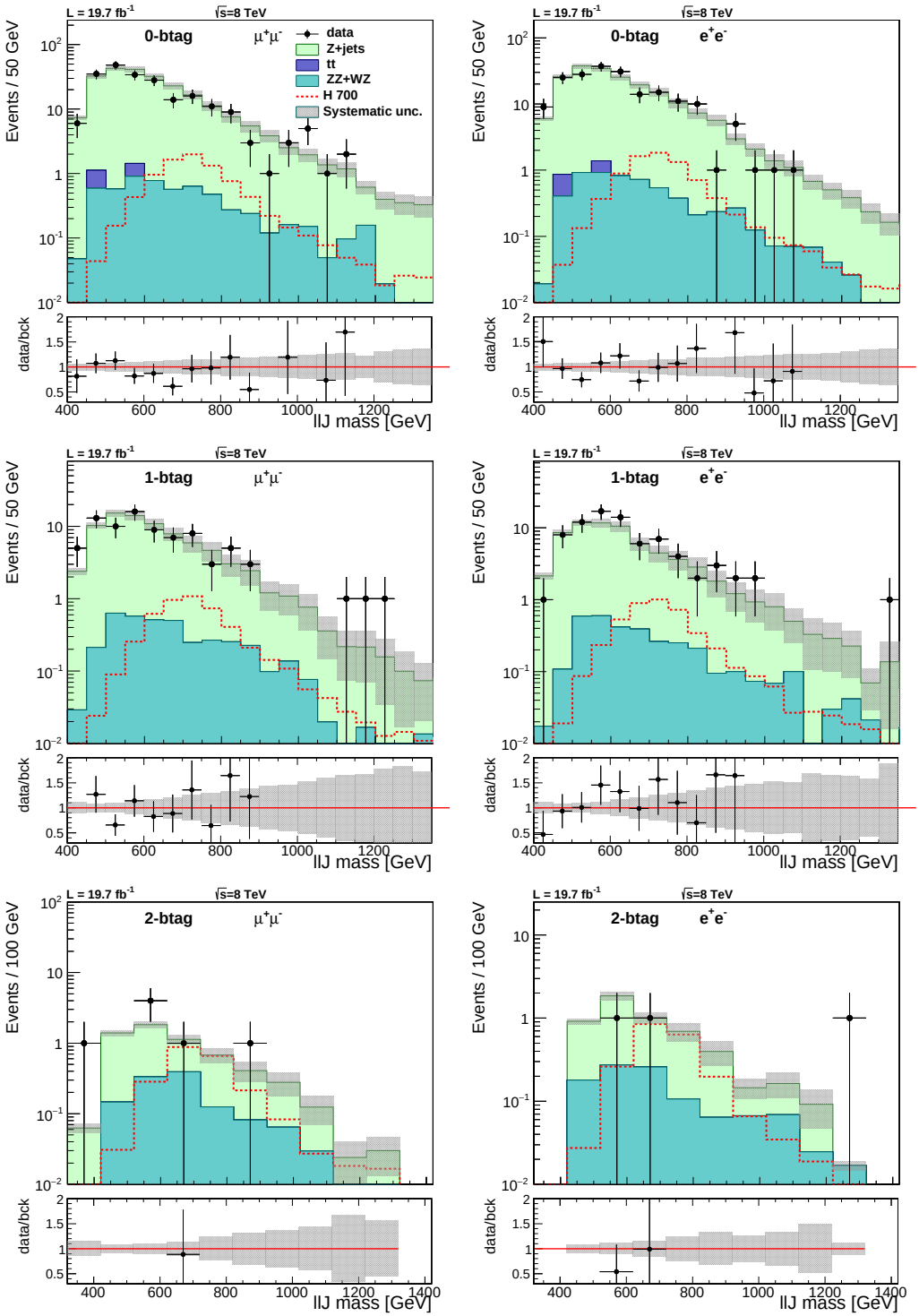


Figure 5.3: Distribution of $m_{\ell\ell q\bar{q}}$ distributions for the merged jet analysis after final selection in the six categories: muon (left column) and electron (right column) channels; 0-btag (top), 1-btag (middle) and 2-btag (bottom). Dots with error bars denote the observed data, the histograms represent the expected backgrounds, and the grey band correspond to the systematic uncertainties. A 700 GeV SM Higgs boson is also shown (red dashed line). The lower panels show the ratio of the observed data and the expected background.

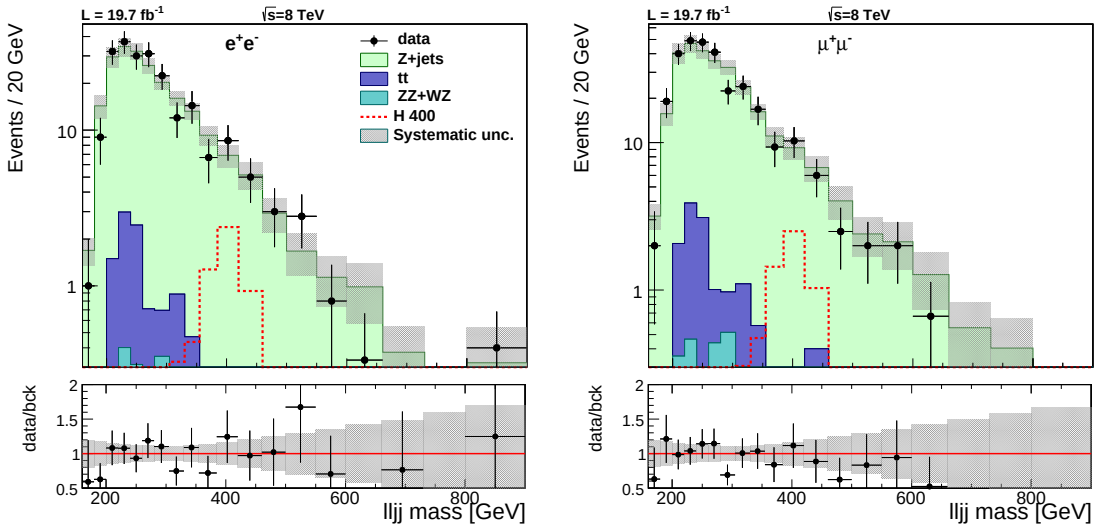


Figure 5.4: Distribution of $m_{\ell\ell q\bar{q}}$ distributions for the VBF analysis after final selection in the electron (left) and muon (right) channels. Dots with error bars denote the observed data, the histograms represent the expected backgrounds, and the grey band correspond to the systematic uncertainties. A 400 GeV SM Higgs boson is also shown (red dashed line). The lower panels show the ratio of the observed data and the expected background.

In order to assess the compatibility of the observed data with the background-only hypothesis or with the presence of a heavy Higgs boson signal, H, a Neaman-Pearson hypothesis testing is performed, the so-called modified frequentist CL_s method [83], using as discriminator the $m_{\ell\ell qq}$ distributions resulting from the analysis. To test the background-only (B) and signal-plus-background (S+B) hypotheses, the binned distributions for B and S+B are fitted to the observed data distributions, using the 14 exclusive categories simultaneously in the calculation.

For each bin i of the $m_{\ell\ell qq}$ distributions, the expected number of events $\langle n_i \rangle$ has contributions from the background and from a hypothetical signal:

$$\langle n_i \rangle = \mu \cdot s_i + b_i \ , \tag{5.1}$$

where b_i and s_i are the expected numbers of events for background and signal, respectively. In this expression, the μ parameter determines the signal strength for a given theoretical model. It is defined as the ratio of the production cross section of the observed signal, $\sigma_{\rm s}$, with respect to the theoretical prediction. For instance, for a SM Higgs boson signal of mass $m_{\rm H}$ the signal strength is:

$$\mu_{\rm SM} = \sigma_{\rm s}/\sigma_{\rm SM}(m_{\rm H}). \tag{5.2}$$

The value $\mu_{\rm SM} = 0$ corresponds to the background-only hypothesis, while $\mu_{\rm SM} = 1$ indicates the presence of a Higgs boson as predicted by the SM. The signal strength

parameter allows a more general interpretation of the observations in terms of models with different cross sections. For BSM theories, the treatment is similar, changing only the signal strength definition to the predicted cross section of the model, σ_{BSM}: $\mu_{\mathrm{BSM}} = \sigma_{\mathrm{s}}/\sigma_{\mathrm{BSM}}(m_{\mathrm{H}})$.

The likelihood that the observed data, with total number of events $\mathrm{N} = \sum n_i$, is compatible with a given hypothesis is the product of the Poisson probabilities for each bin (i):

$$\mathcal{L}(\mathrm{N}|\mu,\theta) = \prod_i \frac{[\mu \cdot s_i(\theta) + b_i(\theta)]^{n_i}}{n_i!} \cdot e^{-(\mu \cdot s_i(\theta) + b_i(\theta))} \ . \tag{5.3}$$

The systematic uncertainties on the shape and normalization of the signal and background predictions are treated as nuisance parameters, θ, taking into account the correlations. To perform the hypothesis testing, the signal strength μ is a free parameter of the fit.

The best test statistic, according to the Neaman-Pearson lemma [84], is defined as the following likelihood ratio, for a given μ hypothesis:

$$Q(\mu) = -2 \cdot ln \frac{\mathcal{L}(\mathrm{N}|\mu,\bar{\theta}_\mu)}{\mathcal{L}(\mathrm{N}|\bar{\mu},\bar{\theta})} \ , \tag{5.4}$$

where $\bar{\mu}, \bar{\theta}$, refer to the values that maximize the likelihood function and $\bar{\theta}_\mu$ are the conditional maximum likelihood estimators of θ, given the μ value. A constraint $0 < \bar{\mu} < \mu$ is applied in this expression: the lower constraint ensures that the signal rate is positive, while the upper condition guarantees that upward fluctuations ($\bar{\mu} > \mu$) are not evidence against the signal hypothesis. In the limit of an infinite number of events, the likelihood ratio is equivalent to a $\Delta\chi^2$ test between the B and μS+B hypotheses. High values of $Q(\mu)$ indicate incompatibility of the data with the signal expectation for a given μ value.

The observed data yield a single value of the test statistic, $Q(\mu_{\mathrm{obs}})$, which is statistically limited. Therefore, two p-values are defined to test the compatibility of the observation with either hypotheses. For μS+B:

$$p(\mu) = p_{\mu s+b} = \mathrm{P}(Q(\mu) \geq Q(\mu_{obs})|\mu\mathrm{S+B}) = \int_{Q(\mu_{obs})}^{\infty} f(Q(\mu)|\mu,\bar{\theta}_\mu) dQ \ . \tag{5.5}$$

For the background-only hypothesis:

$$p(0) = p_b = \mathrm{P}(Q(\mu) \leq Q(\mu_{obs})|B) = \int_{-\infty}^{Q(\mu_{obs})} f(Q(\mu)|0,\bar{\theta}_\mu) dQ \ . \tag{5.6}$$

The probability density functions, f, for both hypotheses are calculated numerically by random generation of pseudo-experiments, following the expected distributions for background and signal, smoothing out the statistical fluctuations present in the predictions. In this particular case, the asymptotic CL_s approximation [85] is used to obtain analytically the probability density functions from the binned distributions.

A convenient way to quantify the compatibility of the observed data with the μS+B hypothesis, for different μ values, is through the ratio of the signal and background p-values, defined as:

$$CL_s(\mu) = \frac{p_{\mu s+b}}{1 - p_b}. \tag{5.7}$$

In the case of an excess of observed data with respect to the B hypothesis, $CL_s(\mu)$ quantifies the statistical significance of the signal. On the other hand, if an overall agreement is observed within the background statistical fluctuations, the CL_s method allows to determine upper limits to the production cross section of a hypothetical Higgs boson signal at a certain confidence level (C.L.). Specifically, the μS+B hypothesis is excluded at the 1-α confidence level when $CL_s < \alpha$. An a-priori value of $\alpha = 0.05$ is chosen, which is equivalent to excluding signals with two standard deviations above the background prediction. The statistical method described in this section is implemented in a CMS specific software package [86].

5.2 Standard Model Higgs boson search

A mass-dependent SM Higgs boson search is performed in the mass range between 230 GeV and 1 TeV. A finite set of simulated m_{H} hypotheses is considered (dots in Figure 4.5). The choice of the mass values is driven by the expected resolution on the measurement of the mass and width of the SM Higgs boson.

The invariant mass distributions of the reconstructed $\ell^+\ell^-\mathrm{q\bar{q}}$ system for observed data, expected background and the different signal hypotheses are used in the statistical analysis, as explained in the previous section. The 14 exclusive categories of the analysis have different signal sensitivities and are combined into a single search by fitting them simultaneously. The control regions of each category, which have a negligible signal contribution, are also used in the fit in order to constrain the relative normalization of the background (appendix A).

Since an overall agreement between the observed data and the background expectations is found, the CL_s method is used to extract upper limits on the production

cross section of SM Higgs bosons. The results are expressed in terms of the signal strength, μ_{SM}, allowing their interpretation in models with similar phenomenologies, but different cross sections.

The 95% C.L. upper limits on μ_{SM} are shown in Figures 5.5, 5.6 and 5.7 for the 14 exclusive categories, for the different m_{H} hypotheses. A SM Higgs boson is excluded if the upper limit lies below $\mu_{\mathrm{SM}} = 1$, displayed as a red line. Similarly, for a model that predicts a production cross section $\sigma = 2\sigma_{\mathrm{SM}}$, a signal would be excluded below $\mu = 2$. As there is no evidence of a SM Higgs boson signal in any of the exclusive categories, 95% C.L. upper limits are set on μ_{SM} for the combination of the 14 channels (Figure 5.8), where the expected upper limits for the different topologies (dijet, merged jet and VBF) are also displayed. The sensitivity of the dijet ggH categories is the highest for mass values up to 500 GeV, rapidly decreasing at higher m_{H} values due to the overlap of the boosted jets. The sensitivity of the VBF analysis is almost constant in the whole m_{H} range and is slightly lower than that of the ggH analysis. The most sensitive categories are those containing two b jets in both the dijet and merged jet analyses, which demonstrates the power of the b-tagging.

The result of the statistical analysis yields an expected sensitivity of the combined analysis to exclude SM Higgs bosons in the m_{H} range between 257 GeV and 712 GeV. The analysis of the observed data excludes a SM Higgs boson at 95% C.L. in the range 285 GeV$<m_{\mathrm{H}}<$777 GeV. These stringent limits on a high mass Higgs boson have a significant weight in the CMS combination of the H$\rightarrow$WW and H$\rightarrow$ZZ final states, as detailed below.

Additionally, for the SM Higgs boson (μ_{SM}=1) the contributions of the different production mechanisms are studied separately. Figure 5.9 shows the 95% C.L. upper limits on the production cross section times branching fraction for the ggH and VBF production modes, permitting to compare the results with the case where one of the two production mechanisms is suppressed. A combined limit in the cross section times branching fraction is calculated assuming the relative ggH and VBF contributions predicted by the SM (Figure 5.10).

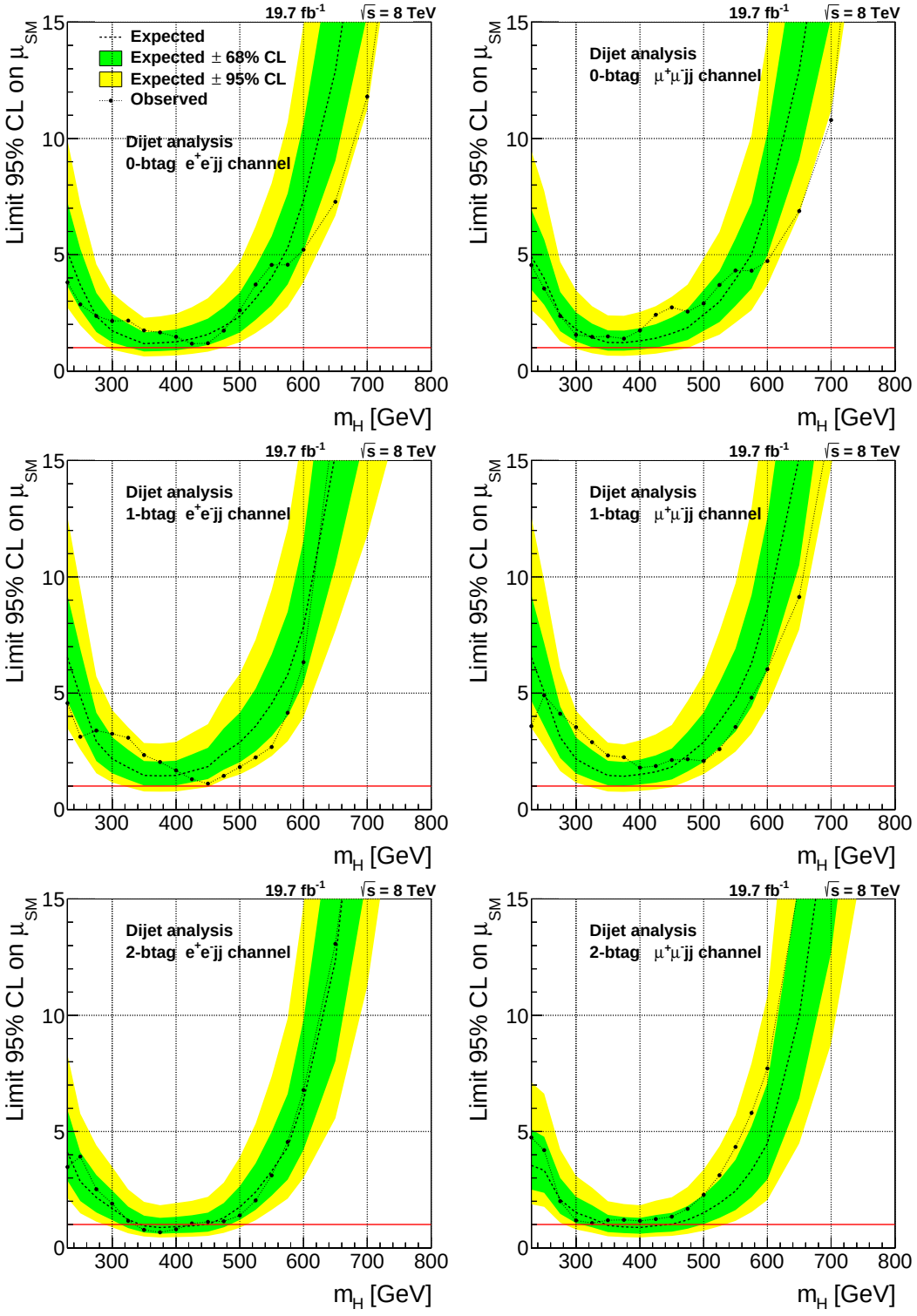


Figure 5.5: Expected (dashed line) and observed (solid line) 95% C.L. upper limits on μ_{SM} as a function of m_{H} for the dijet analysis. Left is for electrons and right for muons. The upper row corresponds to 0-btag, the middle row to 1-btag and the lower row to the 2-btag category.

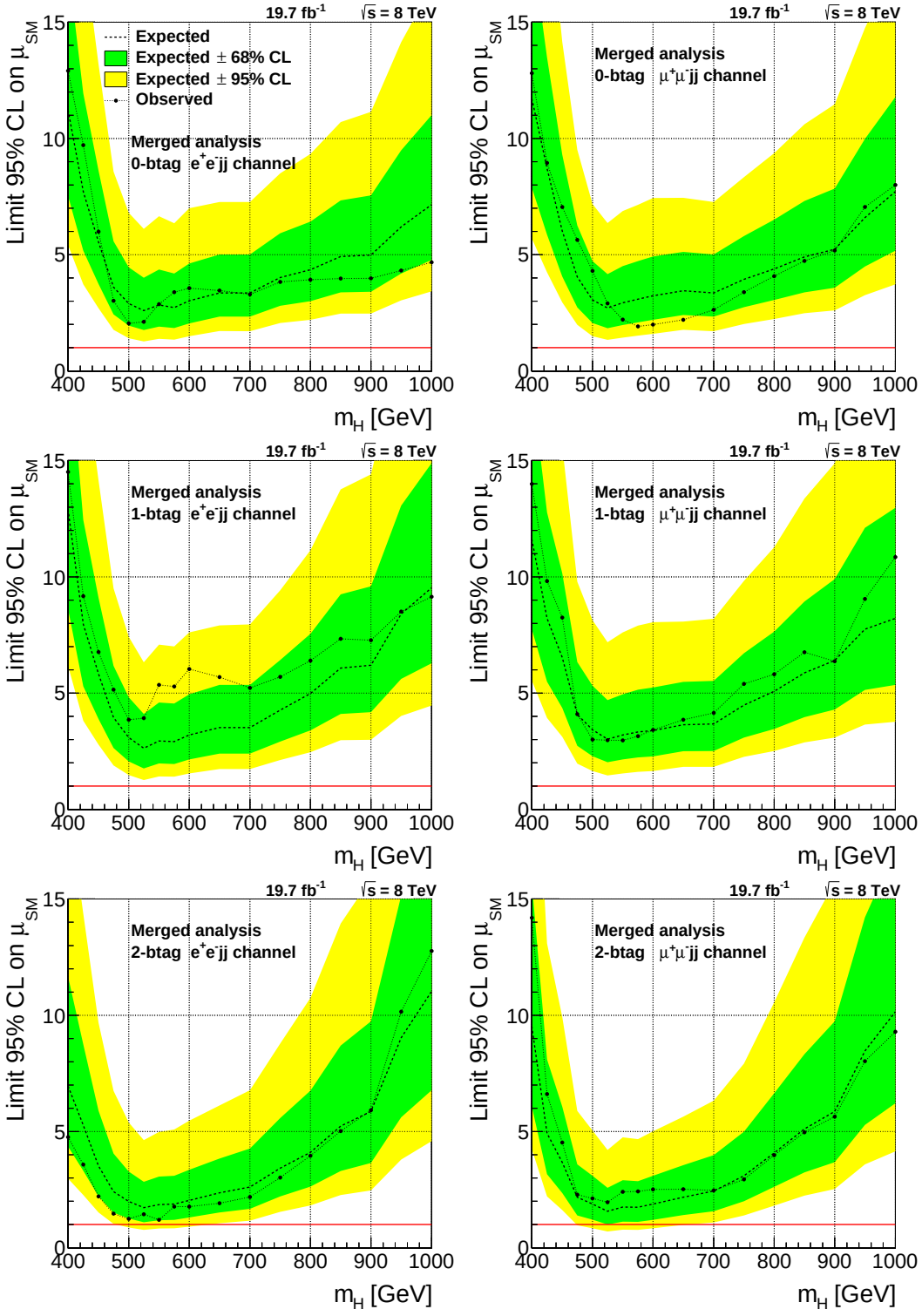


Figure 5.6: Expected (dashed line) and observed (solid line) 95% C.L. upper limits on μ_{SM} as a function of m_{H} for the merged jet analysis. Left is for electrons and right for muons. The upper row corresponds to 0-btag, the middle row to 1-btag and the lower row to the 2-btag category.

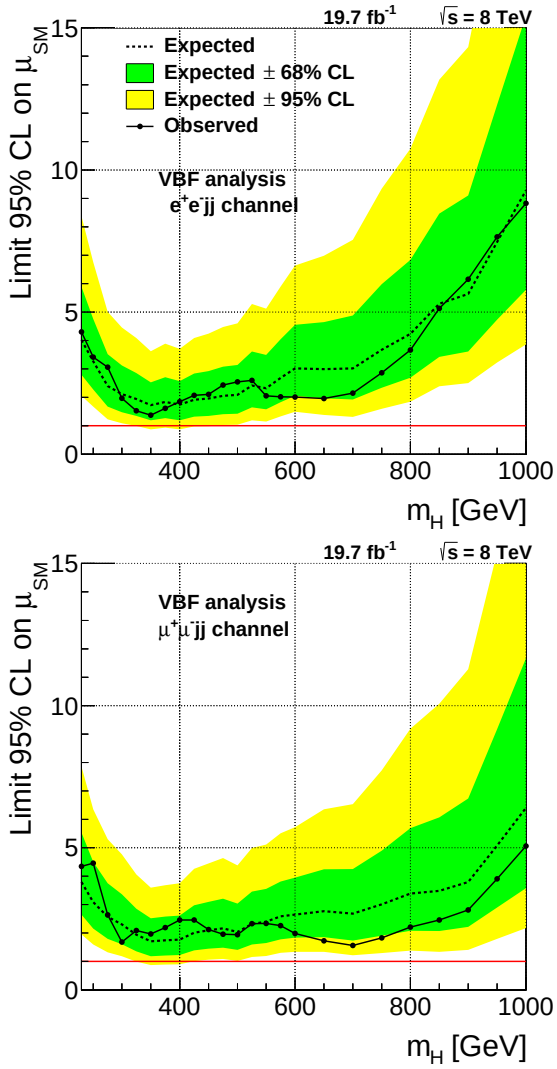


Figure 5.7: Expected (dashed) and observed (solid line) 95% C.L. upper limits on μ_{SM} as a function of the mass for the VBF analysis. The upper plot corresponds to the electron channel and the lower plot to the muon channel.

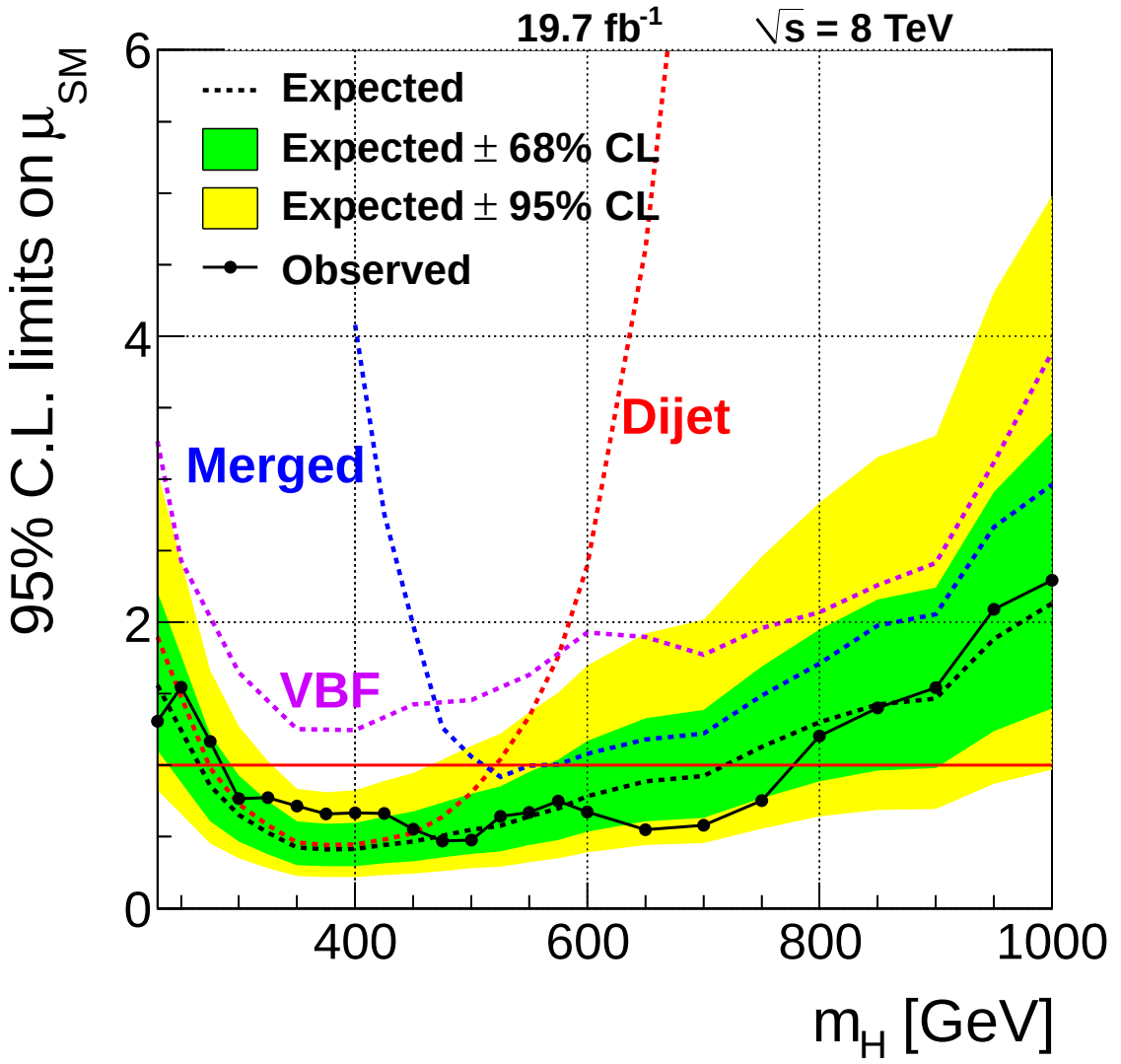


Figure 5.8: Expected (dashed line) and observed (solid line) 95% C.L. upper limits on μ_{SM} as a function of m_{H} for the combined $\mathrm{H} \to \mathrm{ZZ} \to \ell^{+}\ell^{-}\,\mathrm{q\bar{q}}$ analysis. The green (yellow) bands correspond to the 1σ (2σ) uncertainties on the upper limits for the background-only hypothesis. The expected limits for the dijet (dashed red line), merged jet (dashed blue line) and VBF (dashed violet line) sub-analyses are also displayed.

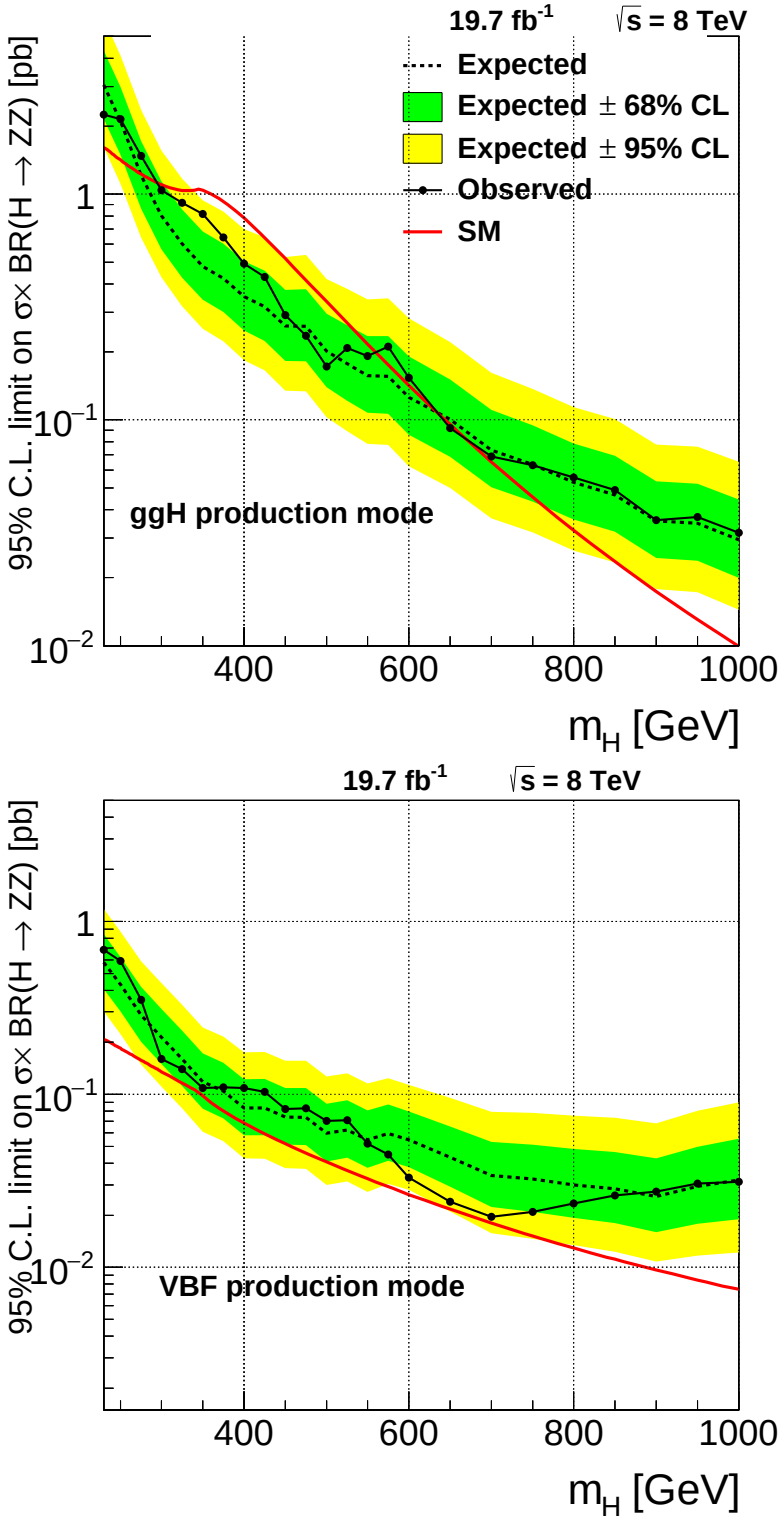


Figure 5.9: Expected (dashed line) and observed (solid line) 95% C.L. upper limits on the total production cross section as a function of m_{H} for the ggH (left) and VBF (right) production mechanisms. The green (yellow) bands correspond to the 1σ (2σ) uncertainties on the upper limits for the background expectation. The solid red line represents the cross section predicted in the SM.

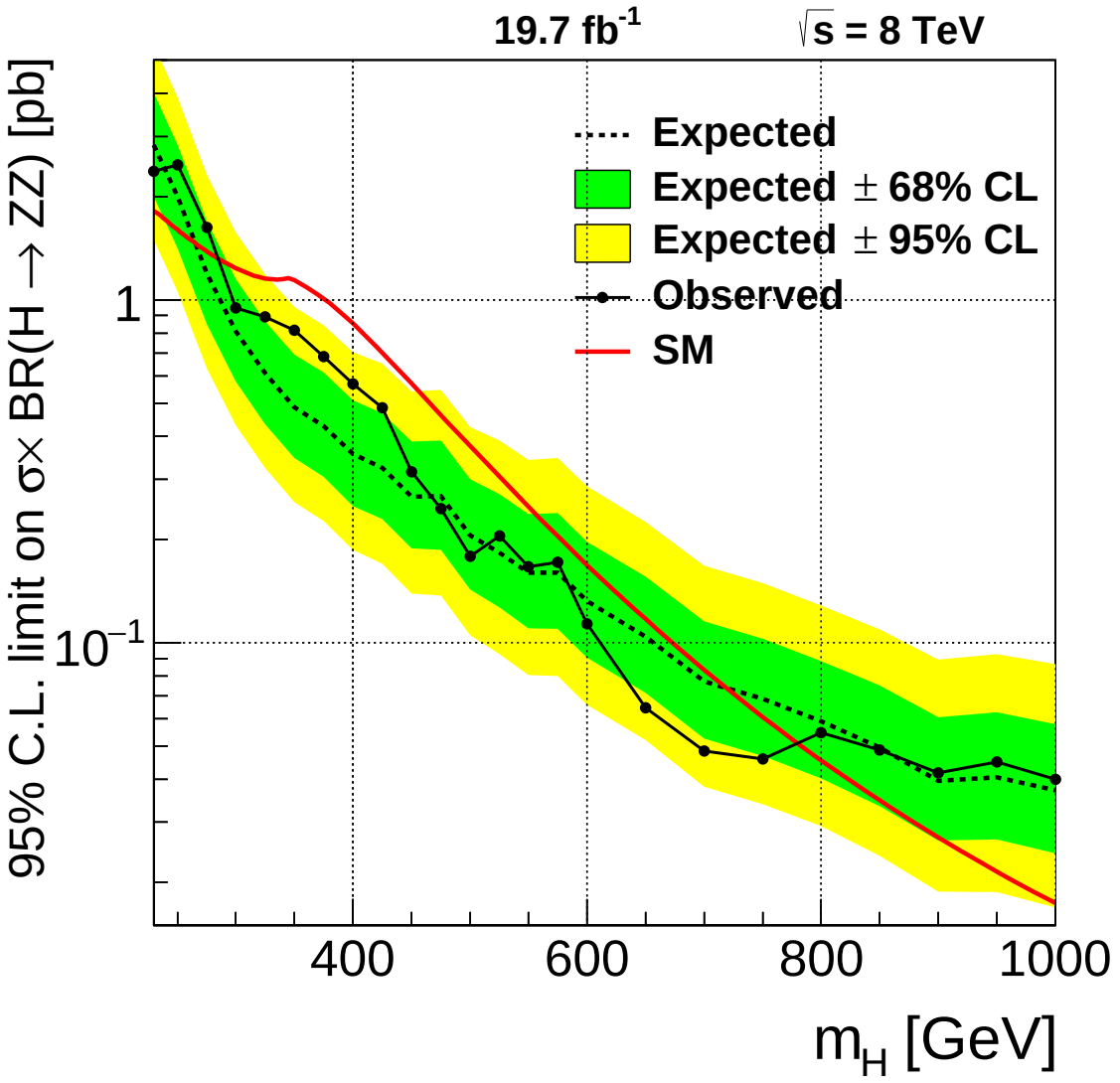


Figure 5.10: Expected (dashed line) and observed (solid line) 95% C.L. upper limits on the total production cross section as a function of m_{H} for the combination of the ggH and VBF production mechanisms assuming the SM cross sections. The green (yellow) bands correspond to the 1σ (2σ) uncertainties on the expectation for the background-only hypothesis. The solid red line shows the cross section predicted in the SM.

H → WW and H → ZZ combined search

The exclusion of additional SM Higgs bosons supports the new boson, h(125), to be the Higgs boson predicted in the SM. To elucidate the existence of extra scalar particles with SM properties in the mass range accessible at the LHC, the CMS collaboration has combined the analyses of the H→W^+W^- and H→ZZ final states in the mass range $m_H \in [145,1000]$ GeV [63]. None of the analyses finds evidence of a signal and, therefore, they are combined into a single analysis to set exclusion limits to the H production. The 95% C.L. upper limits of the combined analysis, along with the observed limits on each of the decay channels considered, is presented in Figure 5.11. The existence of an additional SM Higgs boson is excluded in the whole mass range, up to m_H=1 TeV. The H → ZZ → $\ell^+\ell^-$ q$\bar{\text{q}}$ search presented in this thesis significantly contributes to the combined exclusion at higher masses.

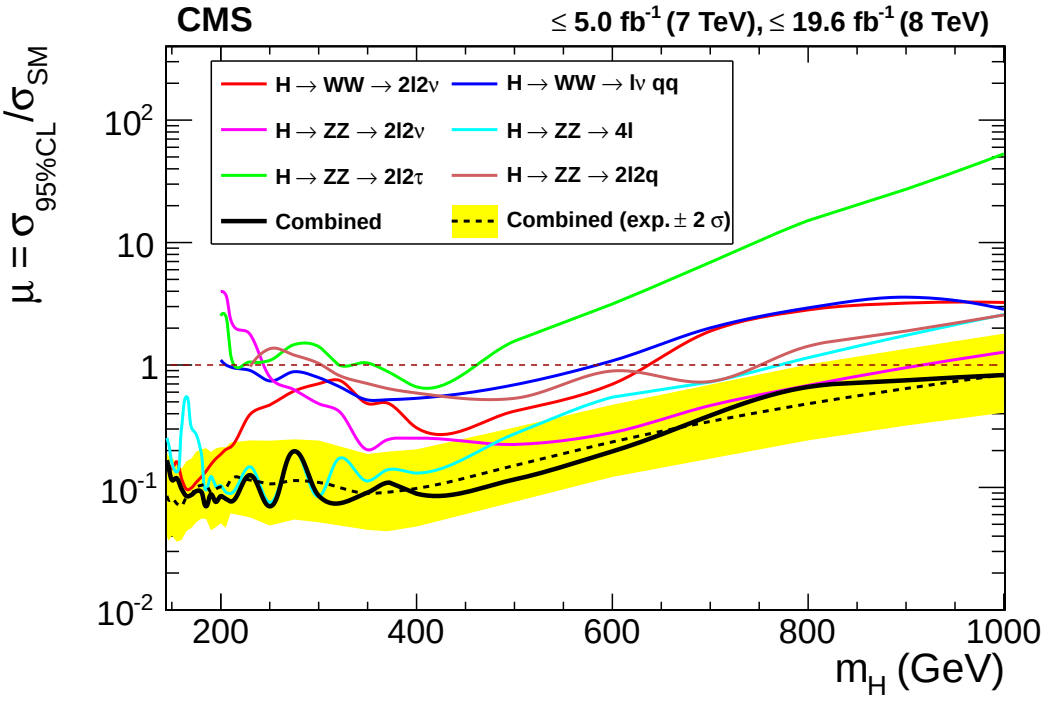


Figure 5.11: Observed 95% C.L. upper limits on the ratio of production cross section over the SM prediction as a function of the mass for the combination of search analyses (black solid line) [63]. The expectation (black dashed line) and its 2σ uncertainty (yellow band) are also shown. The other coloured solid lines depict the observed limits for the individual decay channels.

5.3 Electroweak singlet search

The results of the analysis are also interpreted in the context of the electroweak singlet extension of the SM, which predicts the existence of a scalar particle with properties similar to those of the SM Higgs boson. Other models, such as two-Higgs doublet models (2HDM) or the Minimal Supersymmetric SM (MSSM), will be studied in the coming years by CMS, using the results of the LHC Run II data.

In the electroweak singlet extension, there are two bosons in the scalar sector: the discovered Higgs boson, h(125), and the electroweak singlet particle, H_S, which is mixed with h(125). Three fundamental parameters completely characterize H_S: the mass of the resonance, m_{H_S}, the strength modifier, C', and the branching fraction to decays not predicted in the SM, B_{new}, such as $H_S \rightarrow h(125)h(125)$. These parameters relate the signal strength, μ', and width, Γ', of the H_S resonance with those predicted for the SM Higgs boson:

$$\mu' = C'^2(1 - B_{new})\mu_{SM} \tag{5.8a}$$

$$\Gamma' = \frac{C'^2}{1 - B_{new}}\Gamma_{SM} \ , \tag{5.8b}$$

The strategy developed to search for the SM Higgs boson is used to search for the additional scalar predicted by the electroweak singlet model. The statistical analysis is performed similarly to the SM search, maintaining the background expectations and observed data and weighting the SM Higgs boson predictions to reproduce the signal of the electroweak singlet model using Equations 5.8. Figure 5.12 compares the mass distribution of an electroweak singlet for fixed values of C' and B_{new} with a SM Higgs boson, for a 500 GeV m_H hypothesis. The analysis only considers electroweak singlet particles with equal or smaller widths than the SM Higgs boson, $\Gamma' \leq \Gamma_{SM}$, imposing the condition $C' \leq \sqrt{1 - B_{new}}$.

The statistical analysis is performed on several H_S hypotheses, varying the C' and B_{new} parameters. As an example, the 95% C.L. expected and observed upper limits on the signal strength, μ', are displayed in Figure 5.13, for B_{new}=0 and for several C' values, and in Figure 5.14 for C'^2=0.5 and considering different B_{new} contributions. These results are obtained combining the 14 exclusive categories of the analysis and assuming the relative proportion of ggH and VBF events predicted in the SM.

To fully explore the parameter space of the electroweak singlet model, 95% C.L. upper limits are obtained scanning the (C'^2, m_H) plane, for a fixed value of $B_{new} = 0$, assuming the ggH and VBF relative proportions of the SM (Figure 5.15), and considering both production mechanisms separately (Figure 5.16). The statistical analysis excludes a range of H_S hypotheses with C'^2=1 down to C'^2=0.4 for low m_H values.

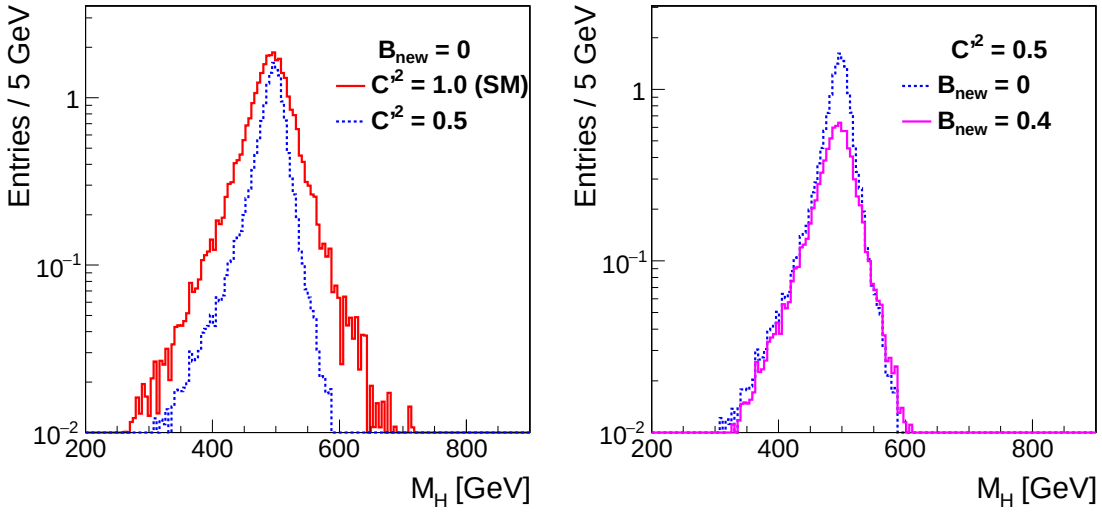


Figure 5.12: Left: mass distributions of a m_{H}=500 GeV SM Higgs boson (solid red line) and an electroweak singlet particle with $C'^2 = 0.5$ and $B_{\mathrm{new}} = 0$ (dashed blue line). Right: mass distributions of a m_{H}=500 GeV electroweak singlet with $C'^2 = 0.5$ for $B_{\mathrm{new}} = 0$ (dashed blue line) and for $B_{\mathrm{new}} = 0.4$ (solid pink line).

Those signals are rather similar to the SM Higgs boson predictions. However, the signals with lower cross sections or very high masses, above 750 GeV, cannot be excluded. The lowest C'^2 value excluded in the analysis is 0.32, corresponding to an electroweak singlet particle of m_{H}=500 GeV. In the case of the separate production mechanisms, the ggH analysis excludes $\mathrm{H_S}$ signals with high C'^2 values up to m_{H}=550 GeV, with the highest sensitivity at m_{H}=500 GeV and C'^2=0.33. The low statistical significance of the sample used in the VBF analysis does not permit to exclude any $\mathrm{H_S}$ signal by itself.

Similarly, to study the possibility of additional decays not predicted for the SM Higgs boson, several scans on the (B_{new}, C'^2) plane are performed, fixing the $\mathrm{H_S}$ mass. Two cases are depicted in Figure 5.17: the most sensitive mass hypothesis, m_{H}=400 GeV, and a very high mass value, m_{H}=900 GeV. The white area, corresponding to $\Gamma' > \Gamma_{\mathrm{SM}}$, is not explored. Only signals in the most sensitive mass range, around 400 GeV, and with properties close to the SM predictions, high C'^2 and low B_{new}, are excluded.

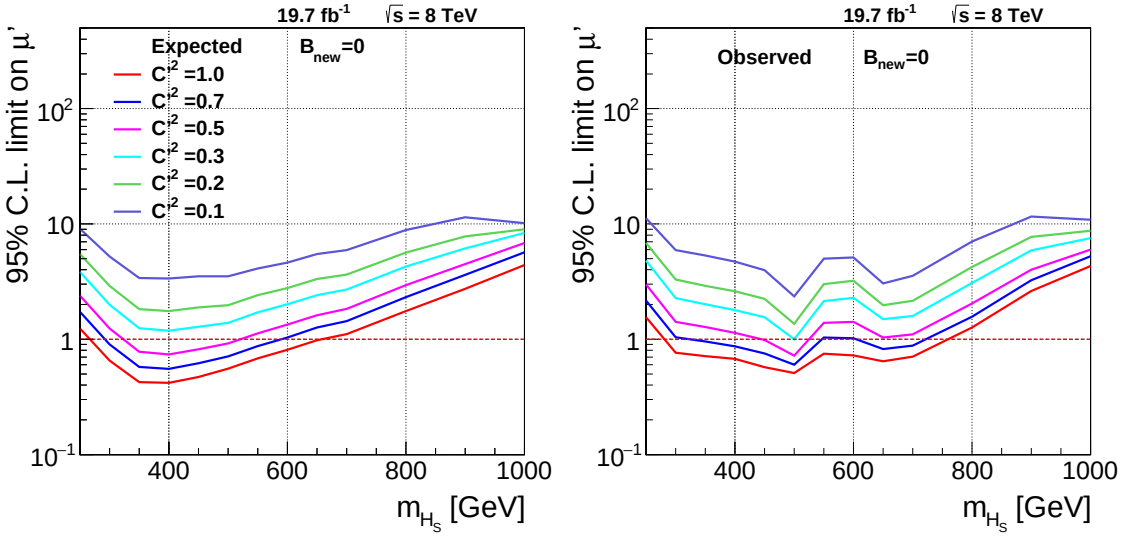


Figure 5.13: Expected (left) and observed (right) 95% C.L. upper limits on μ' as a function of the mass for B_{new}=0. Each line represents a different C'^2 hypotheses.

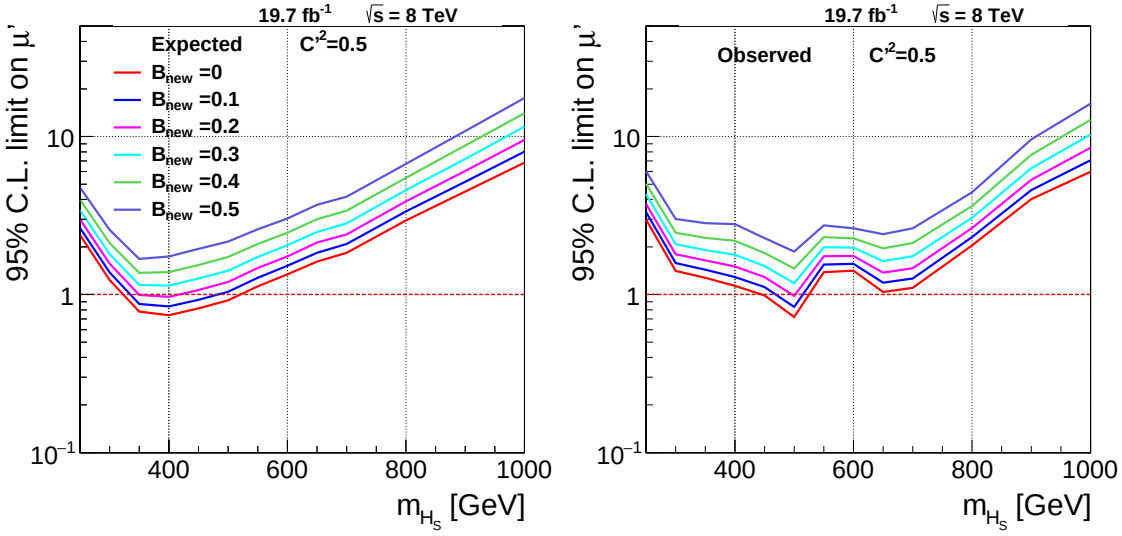


Figure 5.14: Expected (left) and observed (right) 95% C.L. upper limits on μ' as a function of the mass for $C'^2 = 0.5$. Each line represents a different B_{new} hypothesis.

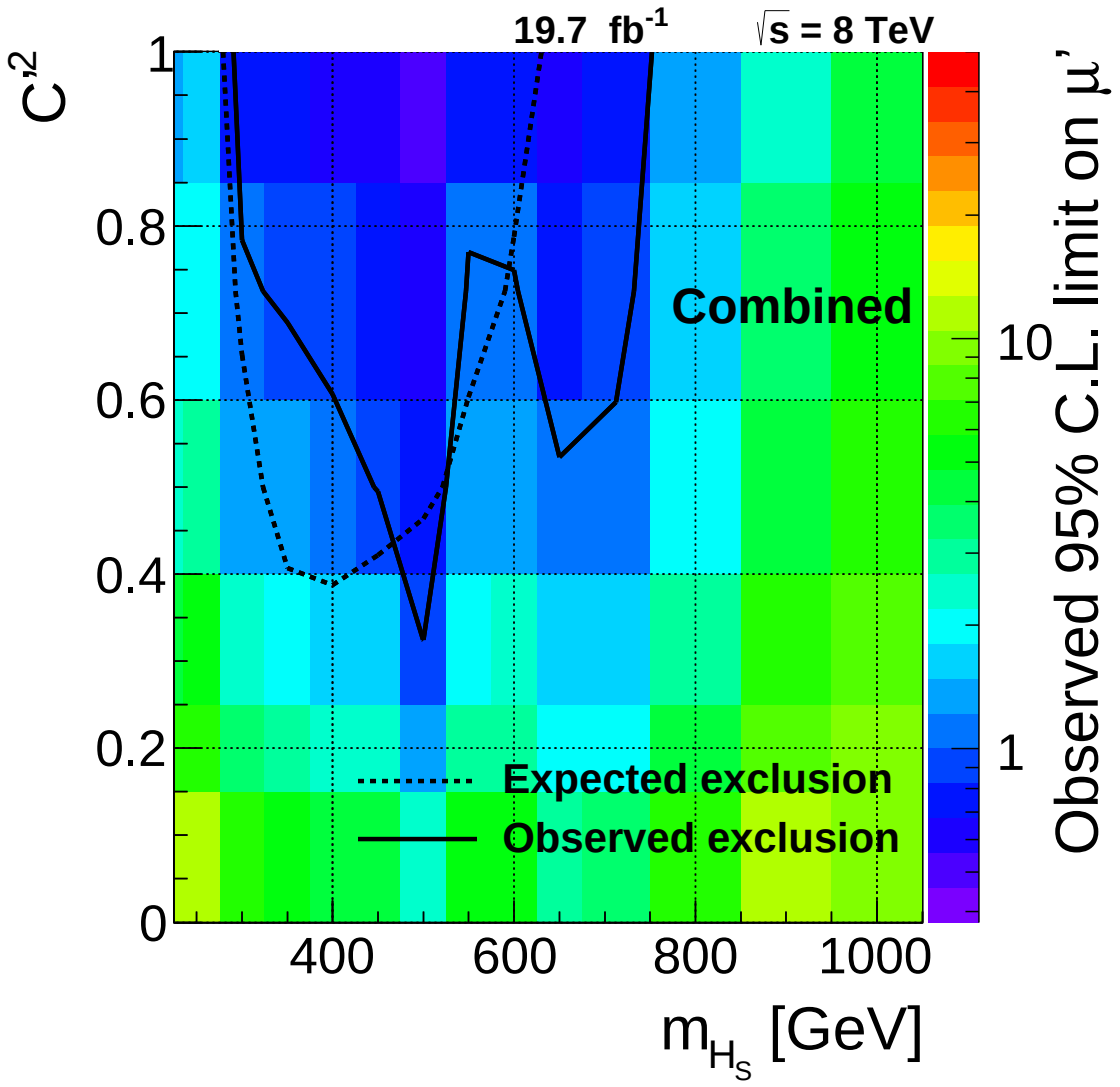


Figure 5.15: Observed limits at the 95% C.L. in the (C'^2, $m_{\mathrm{H_S}}$) plane for B_{new}=0, assuming the ggH and VBF SM rate. The solid (dashed line) lines depict the observed (expected) excluded regions, corresponding to μ'=1.

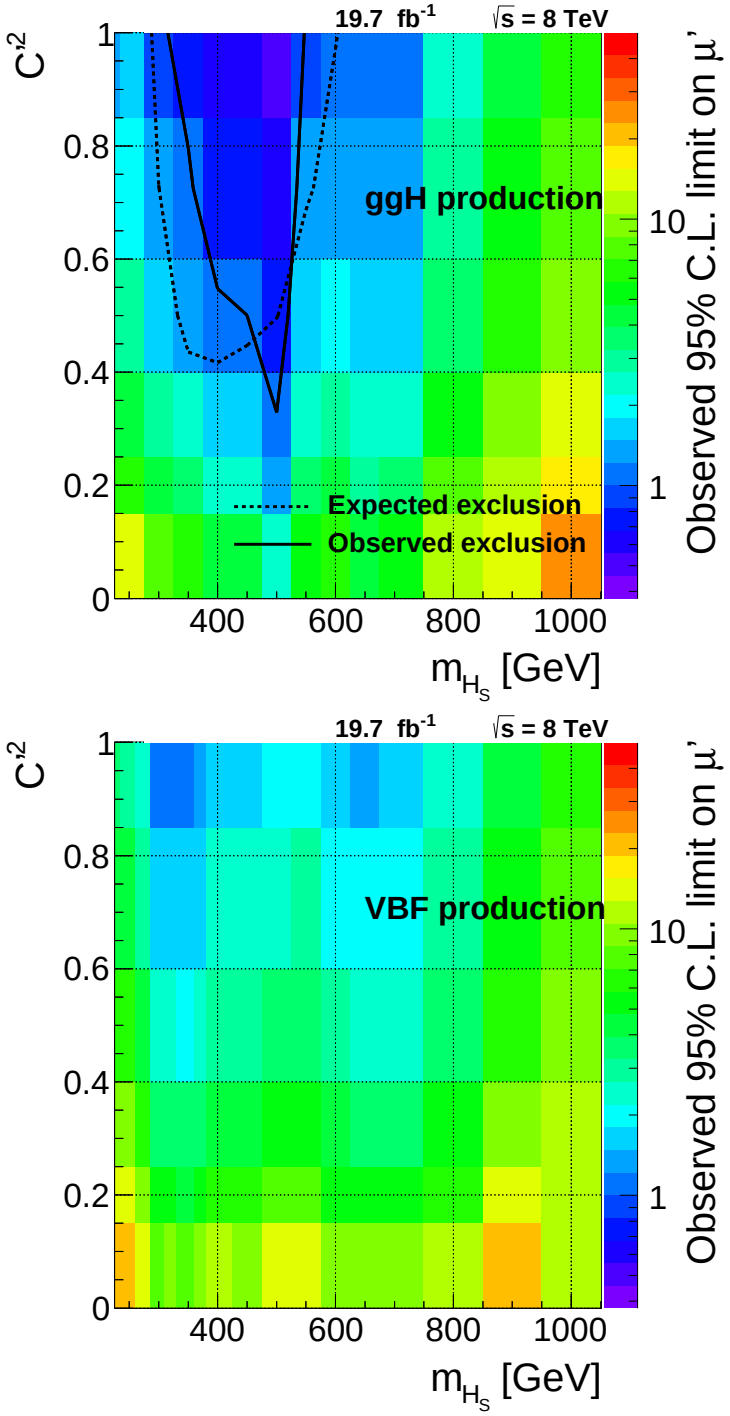


Figure 5.16: Observed limits at the 95% CL in the (C'^2, $m_{\mathrm{H_S}}$) plane for B_{new}=0 for the ggH (upper) and VBF (lower) production mechanisms. The solid (dashed line) lines depict the observed (expected) excluded regions, corresponding to μ'=1.

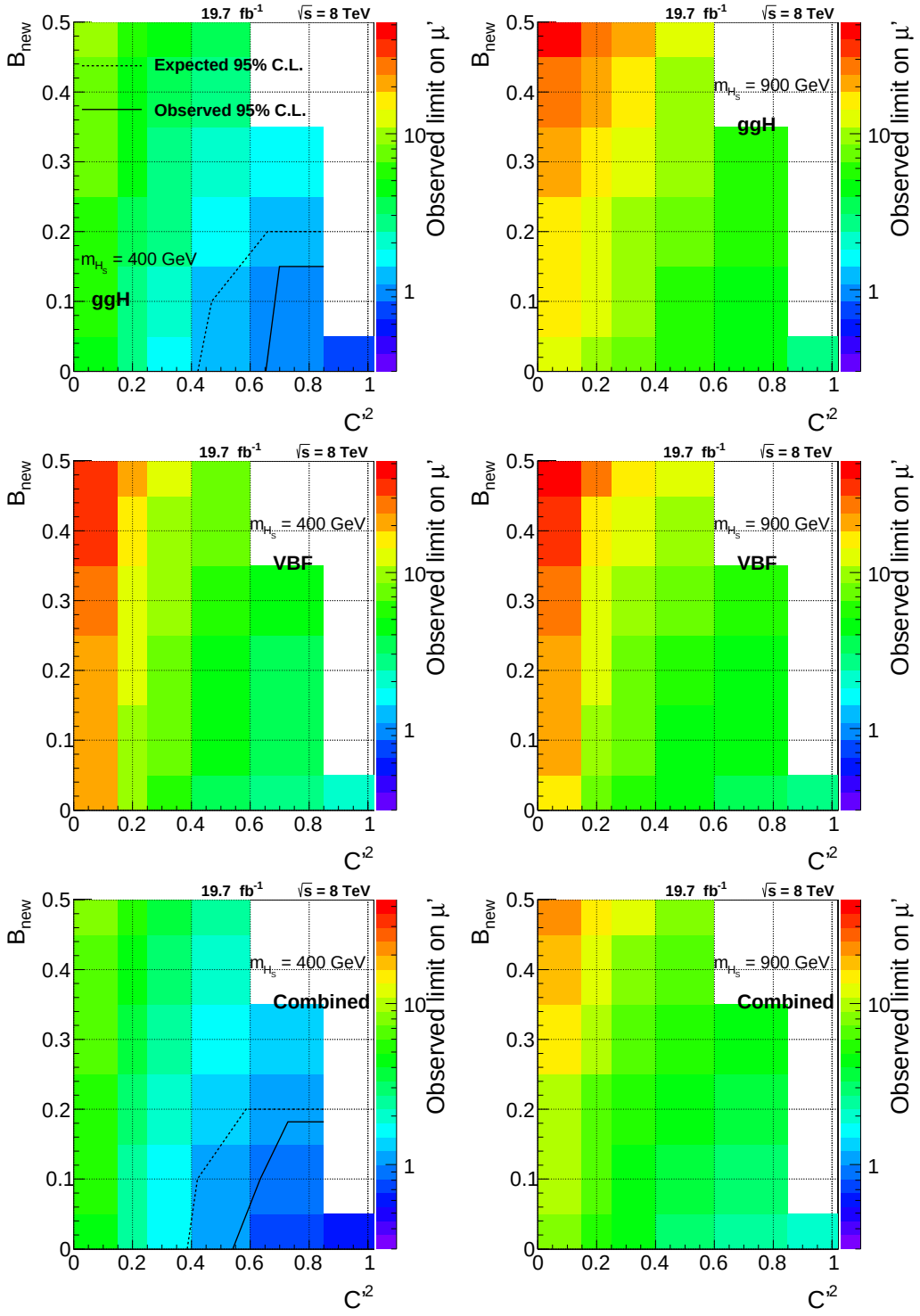


Figure 5.17: Observed limits at 95% C.L. in the (B_{new}, C'^2) plane, for m_{H_S}=400 GeV (left) and m_{H_S}=900 GeV (right). The gluon fusion (upper), VBF (middle) and combined analyses (lower) are depicted. The solid (dashed line) lines represent the observed (expected) exclusion limits, corresponding to μ'=1. The white area represents models not probed in this work.

H → WW and H → ZZ combined search

This analysis is also used in the context of the CMS search for electroweak singlet signals, combining the H→W^+W^- and H→ZZ final states. Since there is no signal evidence in any of these analyses, exclusion limits on the parameter space of the model are set (Figure 5.18). A large region of the (C'^2, m_H) plane is excluded for various values of B_{new}, even below the indirect constrain obtained from the signal strength measurement of h(125), $\mu = 1.00 \pm 0.13$ [26].

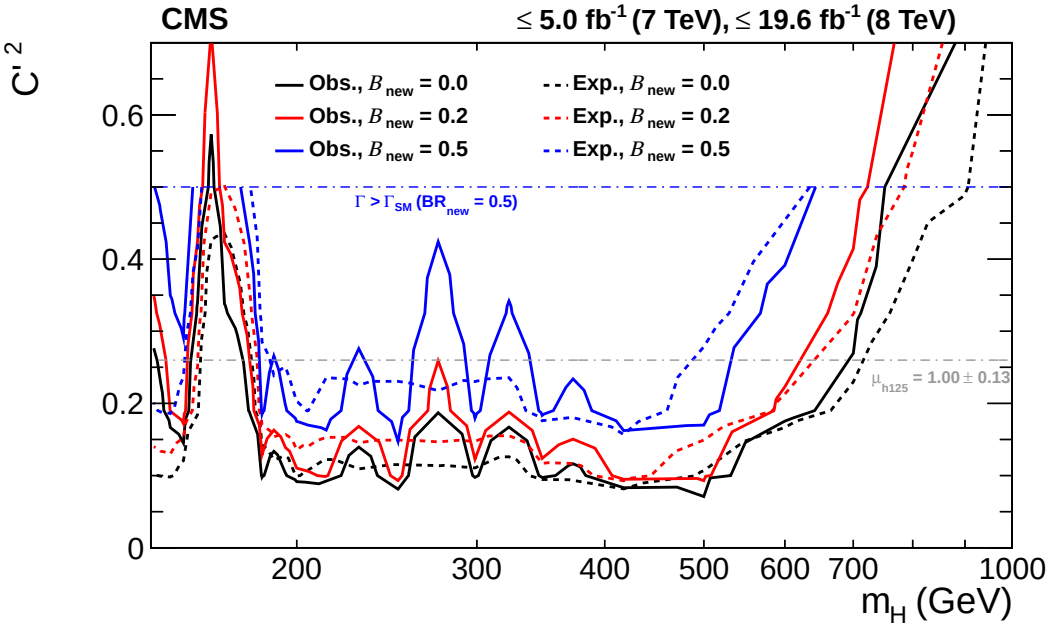


Figure 5.18: Observed (solid line) and expected (dashed line) excluded region at 95% C.L. on the mass, m_H, and strength modifier, C', parameters of the electroweak singlet Higgs boson, combining the ZZ and WW decay channels [63]. Different branching fractions to non-SM decays are tested: $B_{new} = 0$ (black), 0.2 (red) and 0.5 (blue).

◇ ◇ ◇

The LHC has extended the energy frontier of particle physics making possible the discovery of a new scalar particle with the properties of the SM Higgs boson. This thesis has collected the results of the search in the high mass regime, conducted during several years of data gathering with the CMS detector. The decay channel explored, $H \to ZZ \to \ell^+\ell^- q\bar{q}$, made possible to exclude on its own the SM Higgs boson in a broad mass rage, between 285 GeV and 777 GeV. Moreover, the interpretation of the data in the context of the electroweak singlet model extension of the SM has allowed to exclude the existence of an additional Higgs boson in the context of this model in regions of its parameter space. The second three-year run of LHC, just started at the time of this writing, will permit to continue the exploration of the limits of the SM and discard or validate the various theoretically proposed extensions.

Conclusions

Physics results

This thesis is the result of my research in the CMS Collaboration, as a member of the Basic Research Department at CIEMAT (Madrid), started at the end of 2011. I have analyzed proton-proton collisions produced in the LHC at a centre-of-mass energy never reached before in this kind of experiment, 8 TeV, and collected with the CMS detector during the year 2012.

The goal of this work was to search Higgs bosons in the context of the SM and BSM theories, in the decay channel $H \rightarrow ZZ \rightarrow \ell^+\ell^- q\bar{q}$. The analysis explored a wide range of masses of hypothetical scalar particles, between 230 GeV and 1 TeV. The best sensitivity to Higgs boson signals was attained by exploiting the kinematic and topological differences of the events, so that the Higgs boson signal is enhanced and the background suppressed.

A fundamental characteristic of the analysis was the separation of the event sample in exclusive categories, which allowed to increase the overall sensitivity to Higgs bosons. The low-mass region was explored in categories in which two distinct jets can be reconstructed. The high-mass region is populated with events with overlapping jets, merged in a single jet. The capability of efficiently identifying b quarks permitted to increase the discriminating power of the analysis. The analysis was further improved by separating the events according to their production mechanism, either gluon fusion or vector boson fusion.

The ggH dijet analysis was my direct responsibility in CMS and, therefore, the results presented in this work are the official CMS results. In addition, I have significantly contributed to the ggH merged jet and VBF analyses. This thesis presents an improved version of the VBF analysis, in which a higher sensitivity is achieved with simpler selection criteria and analysis techniques. These results are fully compatible with the official CMS results.

In the context of the SM, the statistical analysis confirms the background-only hy-

pothesis, and excludes the existence of a SM Higgs boson in the mass range from 285 GeV to 777 GeV. The analysis of the $\ell^+\ell^-\mathrm{q}\bar{\mathrm{q}}$ Higgs decay channel was part of the CMS effort to cover all possible final states and masses up to 1 TeV, not ruled out by direct searches prior to LHC. Moreover, this result significantly contributes to the confirmation that the measured new scalar state h(125), with mass around 125 GeV, is the Higgs boson predicted in the SM.

The discovery of h(125) with propertis compatible with those of the SM Higgs boson opens the possibility to explore an extended scalar sector. Hence, the SM analysis was expanded to perform a search for scalar particles arising in the electroweak singlet model, which incorporates an additional scalar particle, $\mathrm{H_S}$, mixed with the SM Higgs boson. The result of the electroweak singlet interpretation excludes the presence of a $\mathrm{H_S}$ particle for intermediate masses ($m_{\mathrm{H_S}}$ between 400 and 600 GeV) with high values of C', the strength modifier, and low values of B_{new}, the branching fraction to non-SM decays.

The analysis described in this thesis has been published in two Physics Analysis Summaries [87, 88], public documents that are peer-reviewed by the CMS Collaboration. I have been actively involved in the combination of the H→WW and H→ZZ channels using the full Run I data. The results of this combination [63] that include the work presented in this thesis, have been submitted to the *Journal of High Energy Physics* and will be published imminent. I have presented the results of my research in national and international conferences [89] and in seminars at CIEMAT [90].

Other personal contributions

My responsibilities as a member of the CMS Collaboration included taking Detector Control Shifts (DCS). This task is crucial to ensure that all the sub-systems of CMS are functioning correctly during the data taking period. I also acted as Computing Shift Person (CSP), responsible of monitoring the computing infrastructure for data processing and analysis.

I have carried out additional service work in the CMS collaboration. I have been in charge of studies of muon identification, isolation and trigger, inside the Muon Physics Objects Group (Muon POG), calculating scale factors to correct the simulations. I have been the software validator of the Drift Tube Detector Performance Group (DT-DPG), ensuring that all the software changes involving the reconstruction of muons in the DT chambers were consistent between different releases. Finally, I have done preliminary studies to efficiently trigger top quark events decaying to leptons in the Run II data.

The advanced analysis techniques I have developed for this work within the Higgs Physics Analysis Group (Higgs PAG) will open the possibility to perform searches for heavy scalar particles with the LHC Run II data at $\sqrt{s}$=13 TeV, being collected by CMS since June 2015. The analysis of the $\ell^+\ell^-\mathrm{q}\bar{\mathrm{q}}$ final state with the Run II sample will contribute to the understanding of the Higgs boson physics beyond the SM, probing models such as the 2HDM and different flavours of supersymmetric models like nMSSM and beyond.

Working at the frontier of particle physics at CERN and CIEMAT has been a matchless experience that undoubtedly will shape my future, both as a physicist and as a person.

Appendix A

Distributions in the control region

The $m_{\ell\ell qq}$ distributions in the control region for the 2 VBF categories, the 6 dijet categories and the 6 merged jet categories are displayed in Figures A.1, A.2 and A.3. An excellent agreement, both in the shape and normalization, is observed as expected for a signal free region.

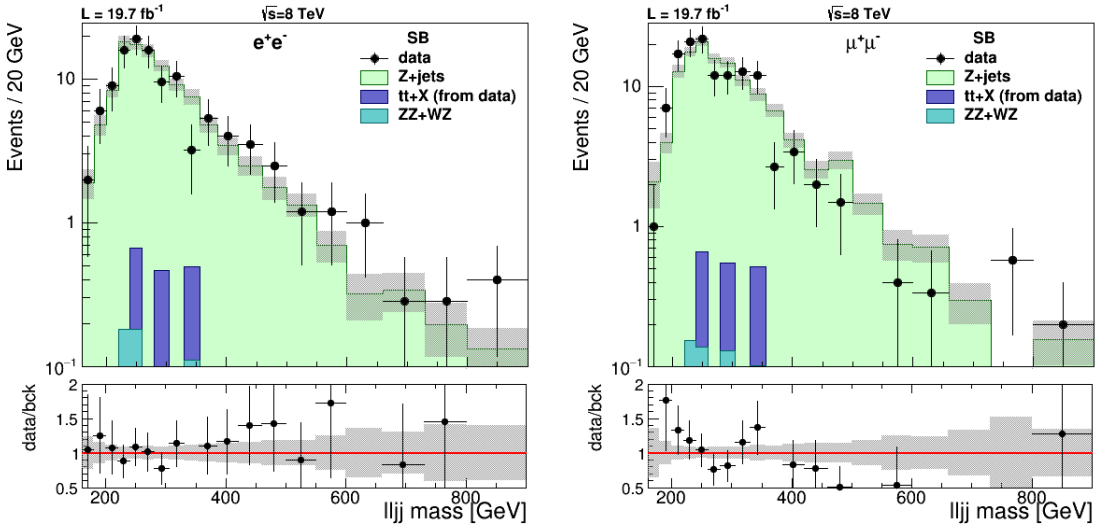


Figure A.1: Distribution of $m_{\ell\ell qq}$ for the VBF analysis after final selection in the control region for the muon (right) and electron (left) categories. Dots with error bars denote the observed data, the histograms represent the expected backgrounds, and the grey band corresponds to the systematic uncertainties. The lower panels show the ratio of the observed data to the expected background.

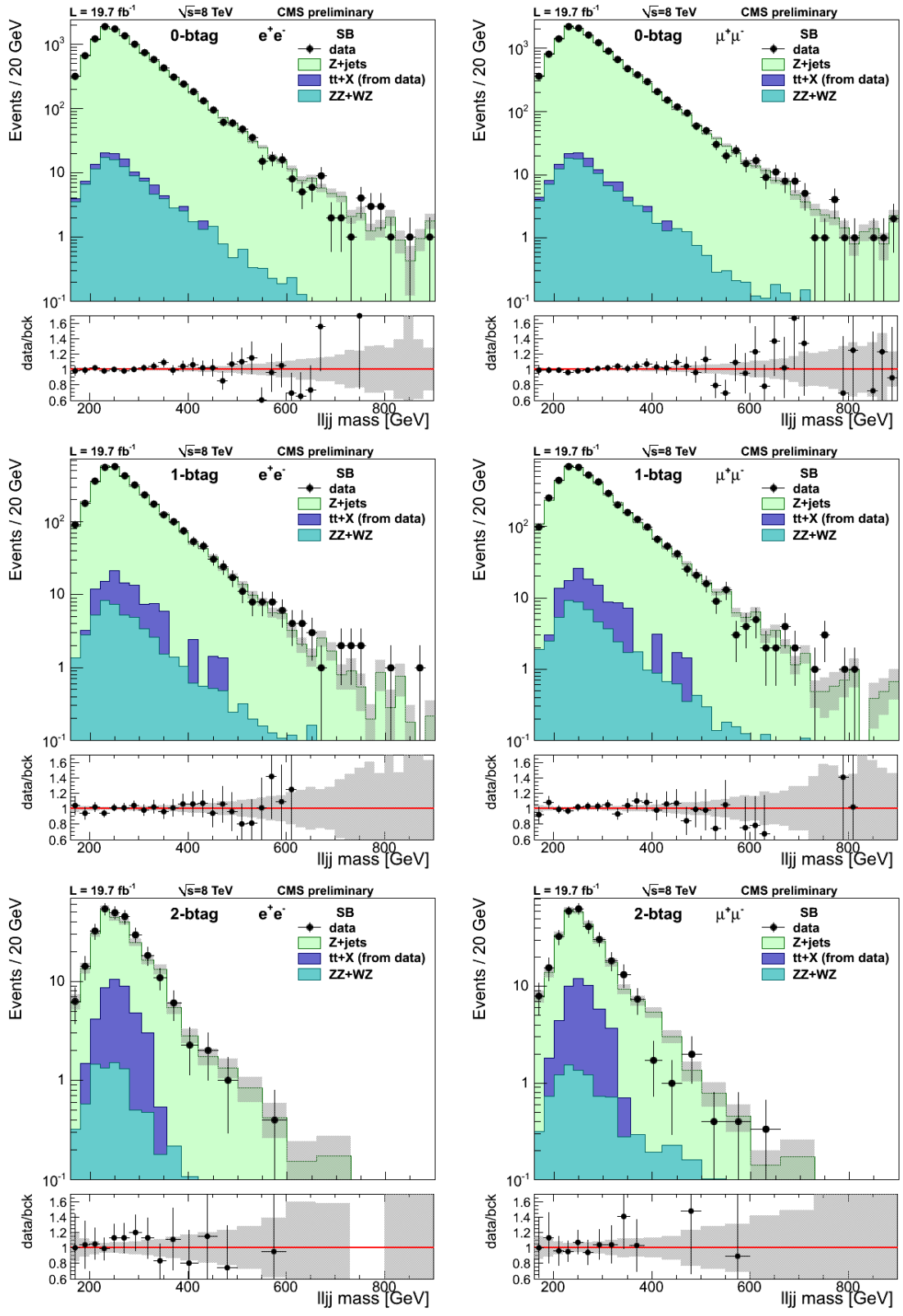


Figure A.2: Distribution of $m_{\ell\ell qq}$ for the dijet analysis after final selection in the control region for the six categories: muon (right column) and electron (left column) channels; 0-btag (top), 1-btag (middle) and 2-btag (bottom). Dots with error bars denote the observed data, the histograms represent the expected backgrounds, and the grey band corresponds to the systematic uncertainties. The lower panels show the ratio of the observed data to the expected background.

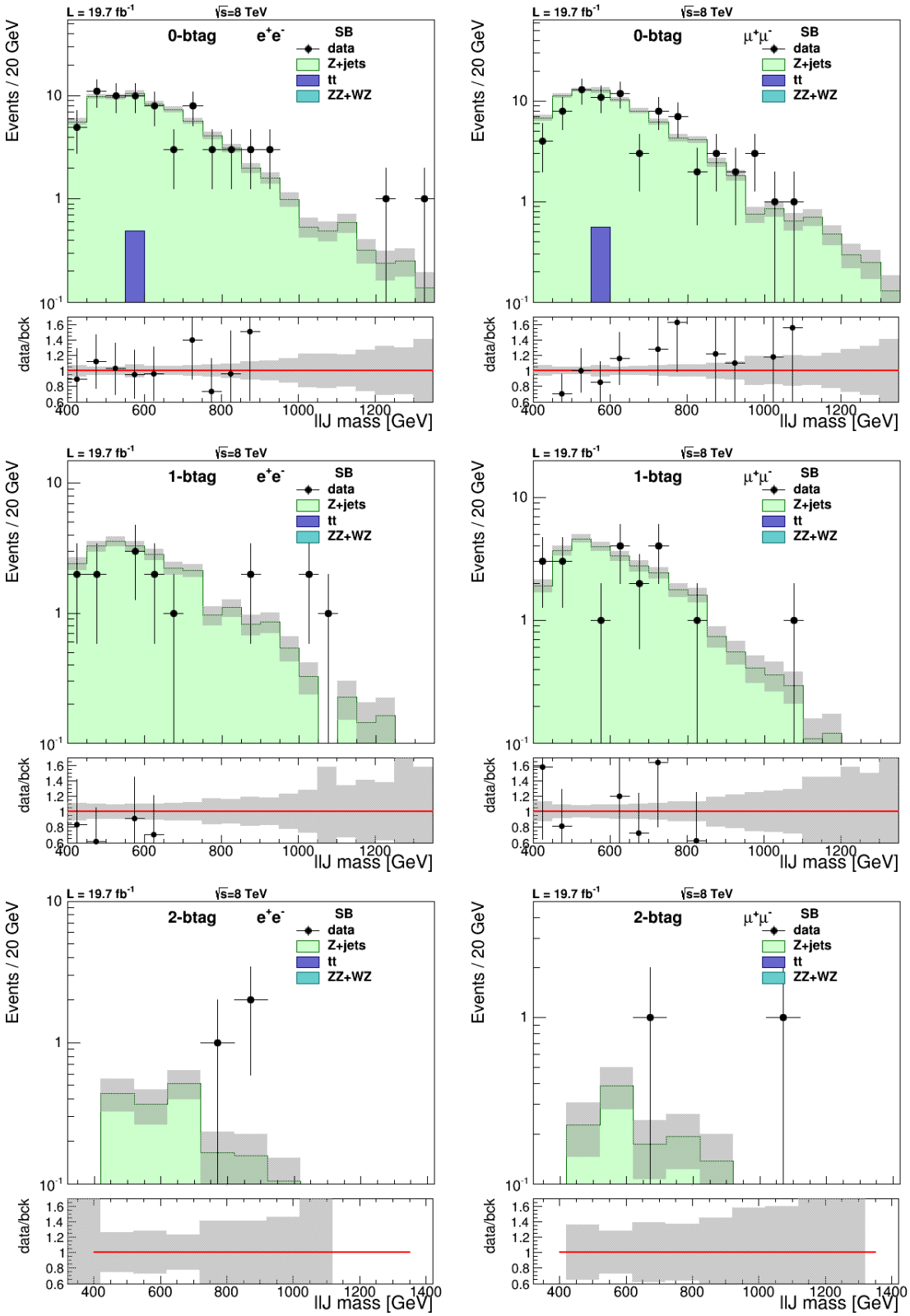


Figure A.3: Distribution of $m_{\ell\ell qq}$ for the merged jet analysis after final selection in the control region for the six categories: muon (right column) and electron (left column) channels; 0-btag (top), 1-btag (middle) and 2-btag (bottom). Dots with error bars denote the observed data, the histograms represent the expected backgrounds, and the grey band corresponds to the systematic uncertainties. The lower panels show the ratio of the observed data to the expected background.

Appendix B

Resumen

B.1 Presentación

El Modelo Estándar de la física de partículas (ME) es una teoría cuántica de campos que describe las partículas elementales y sus interacciones, con una importante excepción, la gravedad. Esta elegante teoría está basada en simetrías gauge fundamentales y permite hacer predicciones de los fenómenos subatómicos con una precisión sin precedentes. Durante los últimos cincuenta años su validez se ha establecido experimentalmente con gran precisión en un amplio rango de energías, hasta la escala electrodébil sobre los 100 GeV.

A pesar de su éxito, se sabe que el ME es una teoría incompleta debido a varias razones. En primer lugar, no describe la gravedad y no incorpora la energía y materia oscuras, firmemente establecidas en las últimas décadas en observaciones cosmológicas. Dichas observaciones revelan que más del 95% del contenido energético del Universo está formado por la materia y energía oscuras. Otras importantes flaquezas del Modelo Estándar son el origen de la masa de las partículas elementales, el problema CP fuerte, las oscilaciones de neutrinos y la asimetría materia-antimateria. El carácter fundamental de todos estos fenómentos motiva fuertemente la necesidad de teorías que extiendan el modelo estándar. Algunas propuestas actuales incluyen la supersimetría, la teoría de cuerdas o teorías con dimensiones adicionales, entre otras.

Las predicciones de dichas teorías se exploran en el Gran Colisionador de Hadrones (LHC) en la escala electrodébil y a escalas energéticas superiores. El LHC colisiona haces de protones a energías sin precedente, de hasta 8 TeV en el centro de masas de la colisión protón-protón hasta el año 2012 y de 13 TeV desde junio de 2015. Este nuevo régimen de energía tiene el potencial de confirmar el ME a la escala de los TeV y de descubrir nuevas partículas asociadas a física desconocida que emerge

a alta energía.

El mecanismo de Higgs, formulado en los años 60, constituía un modelo conveniente para incluir bosones vectoriales masivos en el ME, preservando la renormalizabilidad de la teoría. El colisionador LHC y sus experimentos se diseñaron con el descubrimiento del bosón de Higgs, H, como uno de sus objetivos principales, siendo esta partícula una evidencia directa de la realización del mecanismo de Higgs en la naturaleza. Aunque el ME no proporciona una predicción teórica para la masa del bosón de Higgs, esta se encuentra limitada superiormente a la escala de los TeV, con el fin de evitar la violación de unitariedad de la amplitud de dispersión W^+W^- calculada en teoría de perturbaciones. Antes de la puesta en marcha del LHC, búsquedas directas del bosón de Higgs en experimentos de alta energía (LEP y Tevatron) y límites indirectos obtenidos de las medidas de precisión de parámetros electrodébiles (en LEP, SLD y Tevatron, entre otros) restringen la masa permitida del bosón de Higgs entre 114.4 y 160 GeV y por encima de 170 GeV, con los valores más bajos favorecidos por las medidas indirectas.

El trabajo presentado en esta tesis doctoral consiste en un análisis para la búsqueda de eventos $H\to ZZ\to\ell^+\ell^-q\bar{q}$ utilizando datos recogidos por CMS (Compact Muon Solenoid) en los que los bosones Z se desintegran en dos leptones (muones y electrones) así como en un par quark-antiquark cuyo resultado son dos chorros hadrónicos (llamados *jets*). Este análisis se diseñó originalmente para descubrir un bosón de Higgs del ME en la región de alta masa, hasta 1 TeV. Tras el descubrimiento por los experimentos CMS y ATLAS de una partíucla escalar de masa 125 GeV, compatible con el bosón de Higgs del ME, este análisis ha contribuido a establecer que dicha partícula es la predicha por el ME, excluyendo la existencia de bosones de Higgs adicionales con propiedades del ME en un rango de masas entre 200 GeV y 1 TeV, combinando los resultados con otras búsquedas en canales $H\to WW$ y $H\to ZZ$. En este punto, el análisis se mejoró y pasó a buscar bosones de Higgs de alta masa en teorías más allá del ME (BSM), explorando un modelo con un singlete electrodébil adicional, mezclado con el bosón de Higgs del ME. Al no detectar señales de bosones de Higgs adicionales con el número de eventos recolectados por CMS en el Run I del LHC, este análisis es capaz de excluir amplias regiones del espacio de parámetros del modelo, lo que puede ser fácilmente interpretado como límites en los parámetros de modelos con dos dobletes de Higgs (2HDM).

Las técnicas avanzadas de análisis desarrolladas en esta tesis doctoral servirán en el futuro para buscar señales de bosones de Higgs predichos por teorías BSM en los datos del Run II del LHC, que se están tomando actualmente en el detector CMS a una energía en centro de masas de 13 TeV.

B.2 Conclusiones

Resultados científicos

El trabajo presentado en esta tesis doctoral es el resultado de mi investigación dentro de la colaboración CMS, como miembro del Departamento de Investigación Básica del CIEMAT, que comenzó a finales del año 2011. He analizado colisiones protón-protón producidas en el LHC a una energía en centro de masas nunca alcanzada antes en un experimento de este tipo, 8 TeV, cuyas señales fueron registradas por el detector CMS durante el año 2012.

El objetivo de este trabajo es la búsqueda de un bosones de Higgs en el contexto del ME y de teorías más allá del ME utilizando el canal de desintegración $H \to ZZ \to \ell^+\ell^- q\bar{q}$. El análisis de los datos explora un amplio rango de masas que puede tomar dicha hipotética partícula, entre 230 GeV y 1 TeV. La mayor sensibilidad a señales debidas a bosones de Higgs se alcanza empleando las diferencias cinemáticas y topológicas entre los sucesos, de forma que se suprime la contribución del fondo.

Una característica fundamental del análisis es la separación de la muestra de sucesos en distintas categorías, lo que permite incrementar la sensibilidad a los bosones de Higgs. La región de baja masa se explora en categorías que contienen dos *jets* hadrónicos reconstruidos. La región de alta masa está poblada por sucesos con dos *jets* solapados, que se reconstruyen como un único *jet*. La capacidad de identificar de forma eficiente quarks b (*bottom*) permite incrementar el poder de discriminación del análisis. Separar los sucesos de acuerdo a su mecanismo de producción, fusión de gluones (ggH) o fusión de bosones vectoriales (VBF), también supone una mejora en la búsqueda de bosones de Higgs.

El análisis ggH con dos *jets* separados es mi responsabilidad directa en CMS y, por tanto, los resultados presentados en este trabajo son los resultados oficiales de CMS. Además, he contribuido significativamente a los otros dos análisis, el estudio de ggH con *jets* solapados y el análisis de VBF. Esta tesis presenta una versión mejorada del análisis de VBF, en la que se alcanza una mejor sensibilidad empleando criterios de selección y técnicas de análisis más simples. Mis resultados son perfectamente compatibles con los resultados oficiales de CMS.

En el contexto del ME, el análisis estadístico de los datos confirma la hipótesis del fondo y excluye la existencia de un bosón de Higgs del ME con masas entre 285 GeV y 777 GeV. El análisis del canal de desintegración del bosón de Higgs en $\ell^+\ell^- q\bar{q}$ es parte del esfuerzo común de CMS para cubrir todos los posibles estados finales con

masas hasta 1 TeV, que no habían sido excluidos en experimentos previos al LHC. Es más, este resultado supone una fuerte confirmación de que la nueva partícula escalar h(125), con masa alrededor de 125 GeV, es el bosón de Higgs predicho por el ME.

El descubrimiento de h(125) con propiedades compatibles con el bosón de Higgs del ME abre la posibilidad de explorar un sector escalar extendido con respecto al del ME. El análisis se amplía para realizar una búsqueda de partículas escalares que aparecen en un modelo con un singlete electrodébil, que incorpora una partícula escalar adicional, H_S, mezclada con el bosón de Higgs del ME. El resultado de dicha interpretación excluye la existencia de H_S para masas intermedias (m_{H_S} entre 400 y 600 GeV) con valores altos del parámetro C', la modificación de la fuerza de la senãl, y valores bajos de B_{new}, la fracción de desintegración a partículas no predicha en el ME.

El análisis descrito en esta tesis ha sido publicado como dos *Physics Analysis Summaries* [87, 88], documentos públicos revisados por el Comité de Publicaciones de la Colaboración CMS. He estado activamente involucrado en la combinación de los canales de desintegración H→WW y H→ZZ empleando los datos del Run I. Los resultados de dicha combinación se publicarán próximamente en la revista *Journal of High Energy Physics* [63]. También he presentado los resultados de mi investigación en conferencias nacionales e internacionales [89] y he impartido seminarios en el CIEMAT [90].

Otras contribuciones personales

Mis responsabilidades como miembro de la Colaboración CMS han incluido guardias para supervisar el detector CMS (*Detecor Control Shifts*). Esta tarea es crucial para asegurarse de que todos los subsistemas funcionan correctamente durante la toma de datos. También he realizado guardias del sistema de computación (*Computing Shift Person*), importantes para monitorizar la infraestructura de computación empleada para el procesado y análisis de los datos.

He llevado a cabo otras tareas de servicio para la Colaboración CMS. He estado a cargo de estudios de identificación, aislamiento y *trigger* de muones dentro del grupo *Muon Physics Objects Group*, calculando los factores de escala para corregir las simulaciones. He sido revisor del *software* del grupo *Drift Tube Detector Performance Group*, asegurándome de que los cambios en los algoritmos de reconstrucción de muones en las cámaras DT eran consistentes entre versiones. Finalmente, he realizado un estudio preliminar del *trigger* para el Run II, con objeto de detectar

eficientemente sucesos producidos por quarks top que contienen muones.

Las técnicas avanzadas de análisis que he desarrollado durante mi tesis doctoral en el grupo *Higgs Physics Analysis Group* abrirán la puerta a búsquedas de partículas escalares de alta masa en los datos del Run II del LHC a $\sqrt{s}$=13 TeV, que están siendo recogidos por CMS desde junio de 2015. El análisis del estado final $\ell^+\ell^-\mathrm{q}\bar{\mathrm{q}}$ con esta muestra contribuirá a entender mejor la física del bosón de Higgs en el ME y más allá, considerando modelos como 2HDM y diferentes modelos basados en teorías supersimétricas (nMSSM y más allá).

Trabajar en la frontera de la física de partículas en el CERN y en el CIEMAT supone una experiencia inigualable que, sin ninguna duda, ayudará a dar forma a mi futuro, como físico y como persona.

References

[1] M. Gell-Mann, "A schematic model of baryons and mesons", *Phys. Lett.* **8** (1964), 3, 214 – 215.

[2] D. Gross and F. Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346.

[3] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961), 4, 579–588.

[4] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[5] G. 't Hooft, "Renormalization of Massless Yang-Mills Fields", *Nucl. Phys. B* **33** (1971) 173–199.

[6] UA1 Collaboration, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV", *Phys. Lett. B* **122** (1983) 103–116.

[7] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider", *Phys. Lett. B* **126** (1983), 5, 398–410.

[8] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323.

[9] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133.

[10] P. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509.

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587.

[12] P. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons", *Phys. Rev.* **145** (1966) 1156–1163.

[13] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys. C* **38** (2014) 090001.

[14] ALEPH, DELPHI, L3, OPAL and LEPWG for Higgs Boson Searches Collaboration, "Search for the Standard Model Higgs boson at LEP", *Phys. Lett. B* **565** (2003), 0, 61–75.

[15] H. Flacher et al., "Revisiting the Global Electroweak Fit of the Standard Model and Beyond with Gfitter", *Eur. Phys. J. C* **60** (2009) 543–583.

[16] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *J. High Energy Phys.* **06** (2013) 081–119.

[17] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29.

[18] V. Barger et al., "CERN LHC phenomenology of an extended standard model with a real scalar singlet", *Phys. Rev. D* **77** (2008) 035005.

[19] T. P. Englert et al., "LHC: Standard Higgs and Hidden Higgs", *Phys. Lett. B* **707** (2012) 512–516.

[20] R. Schabinger and J. Wells, "A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the Large Hadron Collider", *Phys. Rev. D* **72** (2005) 093007.

[21] S. Bock et al., "Measuring Hidden Higgs and Strongly-Interacting Higgs Scenarios", *Phys. Lett. B* **694** (2010) 44–53.

[22] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", `arXiv:1307.1347`. CERN-2013-004.

[23] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", `arXiv:1207.7214`. CERN-PH-EP-2012-218.

[24] ATLAS Collaboration, "Evidence for the spin-0 nature of the Higgs boson using ATLAS data", *Phys. Lett. B* **726** (2013) 120–144.

[25] CMS Collaboration, "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs", *Phys. Lett. B* **736** (2014) 64.

[26] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", `arXiv:1412.8662`. Submitted to *Eur. Phys. J. C* (2015).

[27] ATLAS Collaboration, "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC", *Phys. Lett. B* **726** (2013) 88–119.

[28] ATLAS and CMS Collaboration, "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.* **114** (2015) 191803.

[29] O. Bruning et al., "LHC Design Report". CERN, Geneva (2004), CERN-2004-003-v1. http://cds.cern.ch/record/782076.

[30] ATLAS Collaboration, "The ATLAS experiment at the CERN LHC", *J. Instrum.* **3** (2008) S08003.

[31] CMS Collaboration, "The CMS experiment at the CERN LHC", *J. Instrum.* **3** (2008) S08004.

[32] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *J. Instrum.* **3** (2008), S08002.

[33] LHCb Collaboration, "The LHCb experiment at the CERN LHC", *J. Instrum.* **3** (2008), S08005.

[34] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". CERN, Geneva (2006), CERN-LHCC-2006-001. http://cds.cern.ch/record/922757.

[35] CMS Collaboration, "The CMS magnet project: Technical Design Report". CERN, Geneva (1997), CERN-LHCC-97-010. http://cds.cern.ch/record/331056.

[36] CMS Collaboration, "The CMS tracker system project: Technical Design Report". CERN, Geneva (1997), CERN-LHCC-98-006. http://cds.cern.ch/record/368412.

[37] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report". CERN, Geneva (1997), CERN-LHCC-97-033. http://cds.cern.ch/record/349375.

[38] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV", *J. Instrum.* **8** (2013) P09009.

[39] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". CERN, Geneva (1997), CERN-LHCC-97-031. http://cds.cern.ch/record/357153.

[40] CMS Collaboration, "The CMS muon project: Technical Design Report". CERN, Geneva (1997), CERN-LHCC-97-032. http://cds.cern.ch/record/343814.

[41] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV". CERN, Geneva (2010), CMS-PAS-MUO-10-002. http://cds.cern.ch/record/1279140.

[42] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems". CERN, Geneva (2000), CERN-LHCC-2000-38. http://cds.cern.ch/record/706847.

[43] CMS Collaboration, "CMS: The computing project. Technical design report". CERN, Geneva (2005), CERN-LHCC-2005-023. http://cds.cern.ch/record/838359.

[44] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[45] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector". CERN, Geneva (2010), CMS-PAS-PFT-10-001. http://cds.cern.ch/record/1247373.

[46] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". CERN, Geneva (2009), CMS-PAS-PFT-10-001. http://cds.cern.ch/record/1194487.

[47] W. Adam, R. Frhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *J. of Phys. G Nucl. Part. Phys.* **31** 9, N9.

[48] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008), 12, 119–126.

[49] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV with the CMS experiment", *J. High Energy Phys.* **10** (2011) 132.

[50] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *J. High Energy Phys.* **0804** (2008) 063.

[51] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better jet clustering algorithms", *J. High Energy Phys.* **08** (1997) 001.

[52] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *J. Instrum.* **6** (2011) 11002.

[53] CMS Collaboration, "Jet Substructure Algorithms". CERN, Geneva (2011), CMS-PAS-JME-10-013. http://cds.cern.ch/record/1333700.

[54] CMS Collaboration, "Identification techniques for highly boosted W bosons that decay into hadrons", *J. High Energy Phys.* **12** (2014) 17.

[55] S. Ellis, C. Vermilion, and J. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys. Rev. D* **81** (2010) 094023.

[56] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *J. High Energy Phys.* **1103** (2011) 015.

[57] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *J. Instrum.* **8** (2013) P04013.

[58] CMS Collaboration, "Performance of b tagging at $\sqrt{s}$=8 TeV in multijet, tt and boosted topology events". CERN, Geneva (2013), CMS-PAS-BTV-13-001. http://cds.cern.ch/record/1581306.

[59] CMS Collaboration, "Performance of the missing transverse energy reconstruction by the CMS experiment in $\sqrt{s} = 8$ TeV pp data", *J. Instrum.* **10** (2015) P02006.

[60] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *J. Instrum.* **6** (2011) 9001.

[61] B. A. Kniehl and A. Sirlin, "Mass and width of a heavy Higgs boson", *Phys. Lett. B* **440** (1998), 12, 136–140.

[62] Dicus, D.A. and Mathur, V.S., "Upper bounds on the values of masses in unified gauge theories", *Phys. Rev. D* **7** (1973) 3111–3114.

[63] CMS Collaboration, "Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons". Submitted to *J. High Energy Phys.* (2015). arXiv:1504.00936.

[64] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update". CERN, Geneva (2013), CMS-PAS-LUM-13-001. http://cds.cern.ch/record/1598864.

[65] J. Alwall et al., "MadGraph 5 : Going Beyond", *J. High Energy Phys.* **1106** (2011) 128.

[66] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *J. High Energy Phys.* **0605** (2006) 026.

[67] H. Lai et al., "New parton distributions for collider physics", *Phys. Rev. D* **82** (2010) 074024.

[68] S. Kretzer, H. L. Lai, F. Olness, and W. Tung, "CTEQ6 parton distributions with heavy quark mass effects", *Phys. Rev. D* **69** (2004) 114005.

[69] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *J. High Energy Phys.* **11** (2007) 070.

[70] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO vector-boson production matched with shower in POWHEG", *J. High Energy Phys.* **0807** (2008) 060.

[71] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: Multi-Scale Improved NLO", *J. High Energy Phys.* **1210** (2012) 155.

[72] N. Kauer and G. Passarino, "Inadequacy of zero-width approximation for a light Higgs boson signal", *J. High Energy Phys.* **1208** (2012) 116.

[73] S. Goria, G. Passarino, and D. Rosco, "The Higgs Boson Lineshape", *Nucl. Phys. B* **864** (2012) 530–579.

[74] G. Passarino, "Higgs Interference Effects in $gg \to ZZ$ and their Uncertainty", *J. High Energy Phys.* **1208** (2012) 146.

[75] Y. Gao et al., "Spin determination of single-produced resonances at hadron colliders", *Phys. Rev. D* **81** (2010) 075022.

[76] CMS Collaboration, "Pileup Jet Identification". CERN, Geneva (2013), CMS-PAS-JME-13-005. http://cds.cern.ch/record/1581583.

[77] CMS Collaboration, "Identifying Hadronically Decaying Vector Bosons Merged into a Single Jet". CERN, Geneva (2013), CMS-PAS-JME-13-006. http://cds.cern.ch/record/1577417.

[78] P. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev. D* **78** (2008) 013004.

[79] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009), 2, 189–285.

[80] R. D. Ball et al., "Impact of heavy quark masses on parton distributions and LHC phenomenology", *Nucl. Phys. B* **849** (2011), 2, 296 – 363.

[81] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", `arXiv:1101.0538`.

[82] CMS Collaboration, "Measurement of the WW and ZZ production cross sections in pp collisions at 8 TeV", *Phys. Lett. B* **721** (2013), 45, 190 – 211.

[83] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)". CERN, Geneva (2000), CERN-OPEN-2000-205. http://cds.cern.ch/record/451614.

[84] J. Neyman and E. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Phil. Trans. R. Soc. A* (1933), 231, 289–337.

[85] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011), 2,.

[86] L. Moneta et al., "The RooStats Project", *PoS ACAT2010* (2010) 057.

[87] CMS Collaboration, "Search for a standard model like Higgs boson in the decay channel H to ZZ to $\ell^+\ell^- q\bar{q}$ at CMS". CERN, Geneva (2013), CMS-PAS-HIG-12-024. http://cds.cern.ch/record/1564157.

[88] CMS Collaboration, "Search for a standard model like Higgs boson in the H to ZZ to $\ell^+\ell^- q\bar{q}$ decay channel at $\sqrt{s}$=8 TeV". CERN, Geneva (2015), CMS-PAS-HIG-14-007. http://cds.cern.ch/record/2001558.

[89] E. Navarro De Martino, "Search for a high mass SM-like Higgs boson in the H to ZZ to llqq decay channel in CMS", `arXiv:1505.03278`. To appear in the proceedings of the 50th Rencontres de Moriond, Electroweak session.

[90] E. Navarro De Martino, "Búsqueda de partículas tipo bosón de Higgs en el experimento CMS del LHC". Jornadas de investigación y difusión de la Ciencia: Ciclo de Seminarios Jóvenes Investigadores, (2015), ISBN 978-84-7834-734-6, NIPO 721-15-035-1, Depósito Legal M-17056-2015, CIEMAT, Madrid, Spain.